The Thesis Committee for Fei Cao
Certifies that this is the approved version of the following thesis:

A New Method of Data Quality Control in Production Data Using the Capacitance-Resistance Model

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Larry W. Lake

____
Jean-Philippe Nicot

A New Method of Data Quality Control in Production Data Using the Capacitance-Resistance Model

by

Fei Cao, B. E.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

August 2011

Dedication

To my parents and grandparents

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Larry W. Lake for giving me this precious opportunity to work with him. I appreciate his professional knowledge, patience, guidance and encouragement. I have a great time during my studies at UT. Special thanks are due to Dr. Leon Lasdon for his suggestion and support. Thanks are also due to Dr. J.P. Nicot for serving on my thesis committee.

I am grateful to the faculty and staff of the Petroleum and Geosystems Engineering Department at the University of Texas at Austin for being helpful and supportive during my time here. I am also grateful for the association with the members of the Center for Petroleum Asset Risk Management (CPARM) research group here at the University of Texas at Austin. Their help and support are appreciated.

I was supported financially by the Center for Petroleum Asset Risk Management (CPARM) and Bureau of Economic Geology at UT Austin. Special thanks are due to Dr. Susan D. Hovorka and Mr. Emilio Nunez for their help and guidance.

Finally, I am most grateful for the love and support of my parents and grandparents. My grandparents are always my inspiration with their intelligence and wisdom. They encouraged me to pursue my study and research here at UT. This unique experience has opened a world of opportunities for me. I appreciate that they always believe in me and being supportive during the most difficult times in my life. I deeply appreciate their sacrifice during my study here.

Fei Cao

August 2011

Abstract

A New Method of Data Quality Control in Production Data Using the Capacitance-Resistance Model

Fei Cao, M.S.E

The University of Texas at Austin, 2011

Supervisor: Larry W. Lake

Production data are the most abundant data in the field. However, they can often be of poor quality because of undocumented operational problems, or changes in operating conditions, or even recording mistakes (Nobakht et al. 2009). If this poor quality or inconsistency is not recognized as such, it can be misinterpreted as a reservoir issue other than the data quality problem that it is. Thus quality control of production data is a crucial and necessary step that must precede any further interpretation using the production data.

To restore production data, we propose to use the capacitance resistance model (CRM) to realize data reconciliation. CRM is a simple reservoir simulation model that

characterizes the connectivity between injectors and producers using only production and injection rate data. Because the CRM model is based on the continuity equation, it can be used to analyze the production corresponding to the injection signal in the reservoir. The problematic production data are then put into the CRM model directly and the resulting CRM output parameters are used to evaluate what the correct production response would be under current injection scheme. We also make sensitivity analysis based on synthetic fields, which are heterogeneous ideal reservoir models with imposed geology and well features in Eclipse. The aim is to show how bad data could be misleading and the best way to restore the production data.

Using the CRM model itself to control data quality is a novel method to obtain clean production data. We can then apply the new clean production data in reservoir simulators or any other processes where production data quality matters. This data quality control process can help better understand the reservoir, analyze its behavior in a more ensured way and make more reliable decisions.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Problem Statement

Production rates are the most abundant data in the field. However, they can often be of poor quality because of undocumented operational problems (Nobakht et al. 2009), or changes in operating conditions, or even recording mistakes. Problematic production data have been found to skew resulting model parameters in the capacitance and resistance model (CRM). Although the data corruption problem hasn't been thoroughly studied in the oil and gas industry (Wei et al., 2004), it is necessary to develop a new method to control production data quality to avoid misinterpretation of the reservoir behavior. This thesis concentrates on developing a new method of production data reconciliation for the sake of better reservoir management and decision-making.

The capacitance resistance model (CRM) has several characteristics that make it the ideal candidate for this research. The CRM model is based on the continuity equation, which can be used to analyze the production response under the injection signal in the reservoir. The problematic production data are put into the CRM model directly and the resulting CRM output parameters are used to evaluate what the correct production response would be under current injection scheme. The CRM model does not require a geology model, which makes it easy to use for data cleaning. Otherwise, the geological model and reservoir properties must be specified, which make the data reconciliation process much more complicated and time-consuming. The CRM model also enables the possibility of only constructing simple heterogeneous ideal reservoirs in commercial simulators to perform sensitivity analysis.

The algorithm in the CRM model that aims to clean production data is designed to automatically recover contaminated or corrupted production data. Restore production data can be used in reservoir simulators, or any other process, where production data quality matters to help better understand the reservoir.

1.2 The Capacitance-Resistance Model

The CRM model could be traced back to Albertoni and Lake (2002) who first came up with the idea to quantify communication between wells in a reservoir using only production and injection data. In their research, diffusivity filters were used to account for the time lag and attenuation of the changes between injector and producer pairs. Gentil (2005) continued this research. He mainly contributed to explain the physical meaning of the connectivities as functions of reservoir transmissibility as well as incorporat bottom-hole pressure fluctuation terms into the model. Yousef (2006) proposed using a time constant in the model and made significant sensitivity analysis in synthetic fields. Sayarpour (2008) focused on finding an analytical solution of the governing differential equation in the CRM model. Weber (2009) used a more powerful optimization software (GAMS) to solve for the CRM model parameters and came up with different techniques to clean data and reduce model parameters. Today, the CRM research continues to be active and tends to gain more credibility and interests from oil and gas industry.

The CRM model is an input-output model focusing on describing the relationships between injectors and producers by modeling the total fluid production. Only rate data are required to automatically history match the model and obtain a representation of these relationships. Because the model does not require a geological model and reservoir properties, fewer parameters are necessary to specify the model than a traditional reservoir simulator. This enables the CRM model to have lower computational requirements. This character serves the goal of data reconciliation where several attempts need to be made before obtaining the clean production data.

The complete CRM model includes 3 parts:

1. Three CRM models based on different reservoir control volumes
2. A fractional flow model for oil production
3. An optimization model for the net present value of future oil recovery

For the purpose of data quality control, only the CRM models themselves will be discussed in this thesis, which is the first item above.

1.3 Objective of the Thesis

The objectives of this thesis are:

1. Design an algorithm of data quality control by means of applying the CRM model
2. Conduct sensitivity analysis on different sized synthetic fields
3. Analysis factors that affect the data quality control procedure

1.4 Outline of the Thesis

This thesis consists of 5 parts. Chapter 1 describes the motivation of this thesis.

Chapter 2 presents the new method to control data quality using the CRM model. However, the review of the basic theory of capacitance resistance model will be presented first.

Chapter 3 presents the validation and sensitivity analysis of the data quality control (DQC) process. A 5x4 synthetic field with 5 injectors and 4 producers is used in this validation.

Chapter 4 givesa sensitivity analysis in a larger synthetic field with 25 injectors and 16 producers. The reason why we introduce this larger synfield is because there are more producers in the field, which will tend to produce more trustworthy results compared to a small field with only 4 producers.

Chapter 5 gives the summary and conclusion of the discussion in the previous chapters. It also provides recommendations for future work.

Chapter 2 Data Quality Control Algorithm in the CRM Model

In this chapter a new method to control data quality is proposed. The CRM model itself will be used as a data quality control tool. Since this data quality control algorithm is based on the original CRM model, to introduce the new theory, a review of the CRM model will be presented first.

The second part of this chapter provides the detailed algorithm of the data quality control procedure, which explains how and why it works.

2.1 Introduction to the Capacitance-Resistance Model

Traditional numerical simulation requires constructing a geological model together with other information about properties of the reservoir (Weber 2009). The simulation can be computationally expensive especially for a large field with hundreds of wells. Typical reservoir management decisions attempt to combine and balance reservoir complexities and economic factors to determine an optimal scheme to improve hydrocarbon recovery. It is desirable to develop a new simple reservoir model that uses only information measured at wells to obtain a quick solution to field dynamics, and to inform decisions without long preparation and computation.

Based on the need for a simple reservoir simulation model in reservoir management and optimization, the capacitance resistance model (CRM) is proposed. The capacitance resistance model (CRM) uses linear and nonlinear regression techniques to determine model parameters, thus identifying the connectivity between injectors and producers. A nonlinear programming (NLP) optimization algorithm also is applied in the CRM model to optimize future injection rate by maximizing the discounted net profit.

Previous research has shown that CRM provides an excellent fit to total production rates, oil production rates and also suggestions for reservoir management schemes. Last but not least, the history matching process between model and historic data can reflect important information about the reservoir behavior.

2.1.1 CRMT –Tank Model of the Reservoir

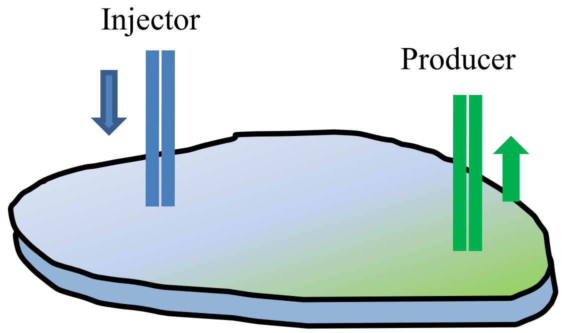


Figure 2.1: Schematic representation of the tank model

Figure 2.1 shows the schematic construction of CRMT, the CRM tank model. The CRMT model has a single inlet and a single outlet. The control volume for the CRM tank model is the drainage volume of the entire reservoir. The CRMT model lumps all the producers into a single pseudo-producer and all the injectors into a single pseudo-injector. Thus, the production and injection rate in the CRMT model are the summation of individual production and injection rate.

The CRM model is based on a continuity equation, which says that the mass of total fluid inside the control volume can only change by the amount that is entering or exiting of the region through the boundary. In the CRMT tank model, the governing continuity equation (Sayarpour, 2008) will be:

$$c_t V_p \frac{d\overline{P}(t)}{dt} = I(t) - q(t) \tag{2.1}$$

where

c_t is the total reservoir compressibility [1/pressure]

V_p is the reservoir pore volume [volume]

$\overline{P}(t)$ is the average reservoir pressure at time t [pressure]

$I(t)$ is the injection rate of the pseudo-injector at time t [volume/time]

$q(t)$ is the total fluid production rate of the pseudo-producer at time t [volume/time].

The definition of productivity index J is (Walsh and Lake, 2003):

$$J = \frac{q(t)}{\overline{P}(t) - P_{wf}(t)} \tag{2.2}$$

where

J is the productivity index[volume/pressure]

$P_{wf}(t)$ is the bottom-hole pressure (BHP) of the pseudo-production well at time t [pressure].

By this definition, the total fluid production rate $q(t)$ can also be expressed as below:

$$q(t) = J(\overline{P}(t) - P_{wf}(t)) \tag{2.3}$$

In equation (2.3), the productive index is assumed to be a constant over the production history. Substituting equation (2.3) into equation (2.1) to eliminate the average reservoir pressure $\overline{P}(t)$ to give:

$$q(t) = I(t) - \frac{c_t V_p}{J}\frac{dq(t)}{dt} - c_t V_p \frac{dP_{wf}(t)}{dt} \tag{2.4}$$

In equation (2.1) and (2.4) above, $I(t)$ should be the effective pseudo injection rate that is contributing to the production. In other words, $I(t)$ doesn't necessarily equal to the summation of injection rates in the field. In the circumstance where injection loss exists, the effective injection rate will be less than total injection rate. On the contrary, the reservoir production could be supported by other driving mechanisms (an aquifer for example). In such cases, the effective injection rate may be greater than the total injection rate. However, we will stick to the notation $I(t)$ to stand for total injection. To distinguish the total injection $I(t)$ and effective injection, one can multiply an injection contribution fraction f to give $fI(t)$ to denote the effective injection, where the fraction f is the fraction of injection that contributes to the production. This fraction f will be referred to gain in this thesis. This fraction f will be less than 1 if injection loss exists and greater than 1 if other support exists. This fraction f will be decided by the CRMT model itself and will be a constant for a specific field. After introducing f (or gain), then the continuity equation (2.4) will give:

$$q(t) = fI(t) - \frac{c_t V_p}{J}\frac{dq(t)}{dt} - c_t V_p \frac{dP_{wf}(t)}{dt} \tag{2.5}$$

Further, we define a "time constant" τ to be:

$$\tau = \frac{c_t V_p}{J} \tag{2.6}$$

Substituting the time constant τ into the continuity equation (2.5) and get:

$$q(t) = fI(t) - \tau \frac{dq(t)}{dt} - J\tau \frac{dP_{wf}(t)}{dt} \tag{2.7}$$

Equation (2.7) is a first-order linear differential equation if τ and J are independent of time. Assume a constant injection rate I_k and linear bottom-hole pressure $P_{wf}(t)$ during a time interval within each time step Δt, one can integrate (2.7) over a discrete time interval Δt:

$$q_k = q_{k-1} e^{-\Delta t / \tau} + (1 - e^{-\Delta t / \tau})(fI_k - J\tau \frac{P_{wf}{}^{k} - P_{wf}{}^{k-1}}{\Delta t}) \tag{2.8}$$

Equation (2.8) is the general working equation for the CRMT model. By replacing q_{k-1} with last time step solution and repeating this process until q_0 which is the production rate at the end of the primary production, a superposition solution in time will be reached as below in equation (2.9):

$$q_k = q_0 e^{-(t_k - t_0)/\tau} + \sum_{n=1}^{k} \left\{ (1 - e^{-\Delta t_n / \tau}) \left[fI_n - J\tau \frac{P_{wf}{}^{n} - P_{wf}{}^{n-1}}{\Delta t_n} \right] e^{-(t_k - t_n)/\tau} \right\} \tag{2.9}$$

where

q_0 is the total fluid production rate at the end of primary production

$P_{wf}{}^{0}$ is the bottom-hole pressure at the end of primary recovery

If bottom-hole pressure data are unavailable (as is often the case), $P_{wf}(t)$ is assumed to be constant over the production history. Thus equation (2.8) and (2.9) will become:

$$q_k = q_{k-1} e^{-\Delta t / \tau} + (1 - e^{-\Delta t / \tau}) fI_k \tag{2.10}$$

$$q_k = q_0 e^{-(t_k - t_0)/\tau} + \sum_{n=1}^{k} \left\{ (1 - e^{-\Delta t_n / \tau}) \left[fI_n \right] e^{-(t_k - t_n)/\tau} \right\} \tag{2.11}$$

The gain and time constant are estimated using a nonlinear regression process. The formulation of the nonlinear program will have the following function as the objective function:

$$\min z = \sum_{k=1}^{n_t} \left(\sum_{j=1}^{n_p} q_{jk}{}^{obs} - q_k \right)^2 \tag{2.12}$$

where

$q_{jk}{}^{obs}$ is the observed total production rate of producer *j* at time step *k*

q_k is the calculated total production rate of the field at time step *k* as defined by either equation (2.8) or (2.10)

n_p is the total number of producers

n_t is the total number of historic time periods.

The gain and time constant are constrained to be positive:

$$\begin{aligned} f &> 0 \\ \tau &> 0 \end{aligned} \tag{2.13}$$

Until now we have formulated a complete nonlinear program optimization problem with an objective function, and a series of constraints as following:

Objective Function: $\min z = \sum_{k=1}^{n_t} \left(\sum_{j=1}^{n_p} q_{jk}{}^{obs} - q_k \right)^2$ (2.12)

Subject to: ① $q_k = q_{k-1} e^{-\Delta t/\tau} + (1 - e^{-\Delta t/\tau})\, fI_k$ (2.10)

Or: $q_k = q_{k-1} e^{-\Delta t/\tau} + (1 - e^{-\Delta t/\tau})(fI_k - J\tau \dfrac{P_{wf}{}^{k} - P_{wf}{}^{k-1}}{\Delta t})$ (2.8)

② $f > 0$, $\tau > 0$ (2.13)

There are only two unknowns in this optimization problem, which are the gain and the time constant. The time constant gives a general idea how much time it takes for the fluid to travel from a injector to a producer in the reservoir.

Gain (or f) less than one means that part of the total field injection does not contribute to total field production. Gain greater than one indicates the existence of other support in the reservoir. And based on whether the gain is greater or less than 1, one will be able to know if this reservoir is balanced or not.

The CRMT, as a fast and simple model, can provide overall information of the reservoir as a whole. The values found by solving the nonlinear program problem above prove to be useful in the subsequent, more complicated CRM models. However, models that could reflect detail information of individual well pair performance will be more beneficial in reservoir management and decision-making.

2.1.2 CRMP – The Producer -Based Representation of the Reservoir

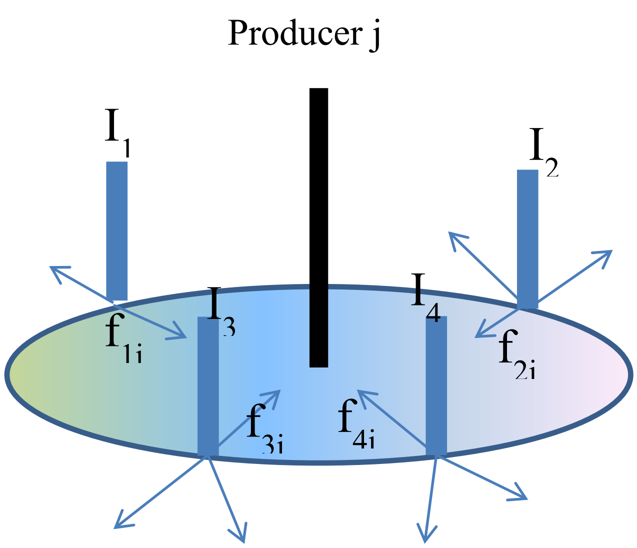


Figure 2.2: Schematic representation of the CRMP model

Figure 2.2 shows the producer-based CRM model (CRMP). For the CRMP, the control volume is the drainage volume around a given producer. The continuity equation for the total fluid production rate of a given production well under such control volume is given by the following equation (Liang et al. 2006):

$$\frac{dq_j(t)}{dt}+\frac{1}{\tau_j}q_j(t)=\frac{1}{\tau_j}\sum_{i=1}^{n_t} f_{ij}I_i(t)-J_j\frac{dP_{wf,j}}{dt} \tag{2.14}$$

where

$q_j(t)$ is the total production rate of producer *j* at time *t* [volume/time]

$I_i(t)$ is the injection rate of injector *i* at time *t* [volume/time]

f_{ij} is the connectivity between injector *i* and producer *j* [dimensionless]

τ_j is the time constant of the drainage volumes of producer *j* [time]

J_j is the productivity index of producer *j* [volume/(time pressure)]

n_i is the total number of injectors

Similar to the CRMT model, the gain f_{ij} represents the effective fraction of water injected in injector *i* that contributes to the total production in producer *j*.

Assuming a constant productivity index J_j for all producers, constant injection rates for all injectors and linearly varied bottom-hole pressure within each time step Δt, we can integrate equation (2.14) over a discrete time Δt period to get (Sayarpour, 2008):

$$q_{jk}=q_{j(k-1)}e^{-\Delta t/\tau_j}+(1-e^{-\Delta t/\tau_j})(\sum_{i=1}^{n_i} f_{ij}I_{ik}-J_j\tau_j\frac{P_{wf,j}{}^{jk}-P_{wf,j}{}^{j(k-1)}}{\Delta t}) \tag{2.15}$$

Mathematically equation (2.15) shows that a weighted average of the last time step production rate and the combined effect of the current rate of effective injection rate from all injectors and the bottom-hole pressure change at that producer.

In equation (2.13), assuming that f_{ij} and τ_j are constants over all time intervals, replacing $q_{j(k-1)}$ with last time step solution $q_{j(k-2)}$ and repeating this procedure until $k=1$, a superposition solution in time will be reached as below in equation (2.16):

$$q_{jk} = q_{j0} e^{-(t_k - t_0)/\tau_j} + \sum_{n=1}^{k} \left\{ e^{-(t_k - t_n)/\tau_j} (1 - e^{-\Delta t_n/\tau_j}) \left[\sum_{i=1}^{n_i} f_{ij} I_i^{\,n} - J_j \tau_j \frac{P_{wf,j}^{\ \ n} - P_{wf,j}^{\ \ n-1}}{\Delta t_n} \right] \right\} \tag{2.16}$$

where

q_{j0} is the total fluid production rate of producer j at the end of the primary production

$P_{wf,j}^{\ \ 0}$ is the bottom-hole pressure of producer j at the end of the primary recovery.

Equation (2.16) will simplify to the CRMT model if there is only one producer and one injector in the field.

When bottom-hole pressure data is unavailable, $P_{wf,j}$ is assumed to be constant over the production history and that bottom-hole pressure term of the model drops out, which gives simpler version of equation (2.15) and equation (2.16):

$$q_{jk} = q_{j(k-1)} e^{-\Delta t/\tau_j} + (1 - e^{-\Delta t/\tau_j}) \sum_{i=1}^{n_i} f_{ij} I_{ik} \tag{2.17}$$

$$q_{jk} = q_{j0} e^{-(t_k - t_0)/\tau_j} + \sum_{n=1}^{k} \left\{ e^{-(t_k - t_n)/\tau_j} (1 - e^{-\Delta t_n/\tau_j}) \sum_{i=1}^{n_i} f_{ij} I_i^{\,n} \right\} \tag{2.18}$$

The gains and time constants are again estimated using multivariable nonlinear regression process. The required objective function is slightly different than that used for the CRMT model:

$$\min z = \sum_{k=1}^{n_t} \sum_{j=1}^{n_p} (q_{jk}^{\ obs} - q_{jk})^2 \tag{2.19}$$

where q_{jk} is the model-calculated total production rate of producer j at time step k as defined by either equation (2.15) or (2.17). This objective function is constrained by equation (2.19) show below as well as some additional constraints.

$$f_{ij}, \tau_j \geq 0 \tag{2.20}$$

In the CRMP model, the gains and time constants are constrained to be positive as shown in equation (2.20). If the CRMT result shows that the sum of the gains for the whole field is equal or less than one, which means some injection may be lost, the constraint for each injector will be:

$$\sum_{j=1}^{n_p} f_{ij} \leq 1 \text{ for all } i \tag{2.21}$$

The constraint (2.21) indicates a material balance on the injected fluid allowing for lost injection within the control volume. The gains in (2.21) are summed over the producer index *j*, which means the application of this constraint requires solving for the model parameters for all producers at the same time. In other words, each producer's model parameter must be found at the same time. However, we often ignore (2.21), so that each producer can be solved independently; this procedure is used to determine initial values for each producer at the very first step of the optimization process.

If the CRMT result shows that gain for the whole field is strictly greater than one, this indicates an aquifer or injection outside the reservoir may be present. In such a case, the constraint (2.21) for each injector will be relaxed, which means the summation of gains $\sum_{j=1}^{n_p} f_{ij}$ of injector *i* can be less, equal or greater than 1. The reason we don't constrain $\sum_{j=1}^{n_p} f_{ij}$ to be greater than 1 under an unbalanced circumstance is because even though the whole field is unbalanced, when it comes to individual wells, the balance situation can vary. Some injectors may have injection loss while others may be supported by other mechanisms (like an aquifer). However, individual variation doesn't conflict with the overall combined unbalance effect.

The CRMP model is a more detailed model that will give an insight into the connectivity of each well pair. The number of parameters and the computational time it takes, though much greater than with the CRMT, are drastically smaller than a traditional reservoir simulator.

2.1.3 CRMIP – The Injector-Producer Pair-Based Representation of the Reservoir

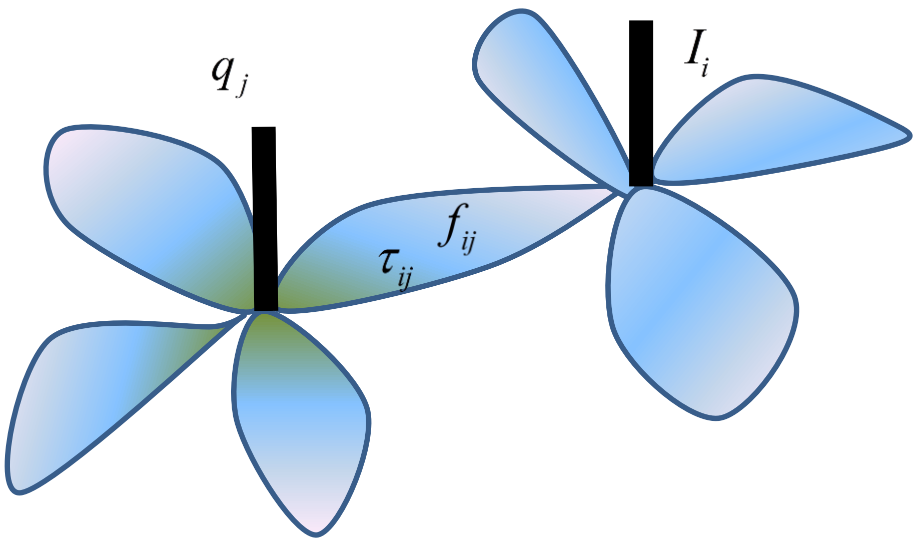


Figure 2.3: Schematic representation of the CRMIP model

The control volume represented in the Figure 2.3 is the drainage volume between an injector-producer pair. The governing continuity equation of the total fluid production for this control volume (Yousef et al. 2006) is:

$$\frac{dq_{ij}(t)}{dt}+\frac{1}{\tau_{ij}}q_{ij}(t)=\frac{1}{\tau_{ij}}\sum_{i=1}^{n_i} f_{ij}I_i(t)-J_{ij}\frac{dp_{wf,j}}{dt} \tag{2.22}$$
$$for\ i=1,2,3...,n_i\ and\ j=1,2,3...,n_j$$

where

$q_{ij}(t)$ is the part of total production in producer *j* that is supported by injector *i* at time *t* [volume/time]

τ_{ij} is the time constant between injector *i* and producer *j* [time]

J_{ij} is the productivity index of the partial production $q_{ij}(t)$ [volume/(time pressure)] defined by equation $q_{ij}(t)=J_{ij}(\overline{P}_{ij}(t)-P_{wf,j}(t))$, where $\overline{P}_{ij}(t)$ is the average reservoir pressure between injector *i* and producer *j*.

Summing equation (2.22) over all injectors will give the production rate of producer j in a multiwell system (Youself, 2006):

$$q_j(t)=-\sum_{i=1}^{n_i}\tau_{ij}\frac{dq_{ij}(t)}{dt}+\sum_{i=1}^{n_i} f_{ij}i_i(t)-\frac{dP_{wf,j}}{dt}\sum_{i=1}^{n_i}\tau_{ij}J_{ij} \tag{2.23}$$

Similar to the CRMP model, equation (2.23) shows that the current total production rate of producer *j* is a weighted average of the primary production rate and the combined effect of injection rate and the bottom-hole pressure change at the well. Note that in equation (2.23) the time constant not only appears in the bottom-hole pressure term, but also in the primary production term. To simplify the above equation, it has been assumed that the time constant τ_{ij} is the same (Youself, 2006) for both primary production and the bottom-hole pressure terms.

The continuity equation (2.23) is integrated analytically over the discrete time period Δt (Sayarpour, 2008) assuming a constant productivity index J_{ij} for all producers, constant injection rates and linearly varying bottom-hole pressures within each discrete Δt to give:

$$q_{ijk} = q_{ij(k-1)} e^{-\Delta t / \tau_{ij}} + (1 - e^{-\Delta t / \tau_{ij}})(f_{ij} I_{ik} - J_{ij} \tau_{ij} \frac{P_{wf}^{\ jk} - P_{wf}^{\ j(k-1)}}{\Delta t}) \tag{2.24}$$

Summing equation (2.24) over the injector index, *i*, to get the total production rate in producer *j* in equation (2.25):

$$q_{jk} = \sum_{i=1}^{n_i} q_{ijk} = \sum_{i=1}^{n_i} \left[q_{ij(k-1)} e^{-\Delta t / \tau_{ij}} + (1 - e^{-\Delta t / \tau_{ij}})(f_{ij} I_{ik} - J_{ij} \tau_{ij} \frac{P_{wf}^{\ jk} - P_{wf}^{\ j(k-1)}}{\Delta t}) \right] \tag{2.25}$$

Now the total fluid production rate for each producer j is the sum of a series of weighted sums of the fluid production associated with each injector-producer pair. In equation (2.25), assuming that f_{ij} and τ_j are constants over all time intervals, one can replace $q_{ij(k-1)}$ with last time step solution $q_{ij(k-2)}$ and repeat this procedure until t_0. The resulting solution is a superposition equation in time from time t_0 in equation (2.26) (Sayarpour, 2008):

$$q_{ijk} = q_{ij}(t_0) e^{-(t_k - t_0) / \tau_{ij}} + \sum_{n=1}^{k} \left\{ (1 - e^{-\Delta t_n / \tau_{ij}})(f_{ij} I^{(n)}{}_i - J_{ij} \tau_{ij} \frac{P_{wf}^{\ jn} - P_{wf}^{\ j(n-1)}}{\Delta t_n}) e^{-(t_k - t_n) / \tau_{ij}} \right\} \tag{2.26}$$

Summing over injector index in (2.26) to get equation (2.27) (Sayarpour, 2008):

$$\begin{aligned} q_{jk} = \sum_{i=1}^{n_i} q_{ijk} &= \sum_{i=1}^{n_i} q_{ij}(t_0) e^{-(t_k - t_0) / \tau_{ij}} \\ &+ \sum_{i=1}^{n_i} \left\{ \sum_{n=1}^{k} \left\{ (1 - e^{-\Delta t_n / \tau_{ij}})(f_{ij} I^{(n)}{}_i - J_{ij} \tau_{ij} \frac{P_{wf}^{\ jn} - P_{wf}^{\ j(n-1)}}{\Delta t_n}) e^{-(t_k - t_n) / \tau_{ij}} \right\} \right\} \end{aligned} \tag{2.27}$$

In absence of bottom-hole pressure data, P_{wf} is assumed to be constant over the production history to give a simpler version of equation (2.25) and (2.27):

$$q_{jk} = \sum_{i=1}^{n_i} q_{ijk} = \sum_{i=1}^{n_i} \left[q_{ij(k-1)} e^{-\Delta t / \tau_{ij}} + (1 - e^{-\Delta t / \tau_{ij}}) f_{ij} I_{ik} \right] \tag{2.28}$$

$$q_{jk} = \sum_{i=1}^{n_i} q_{ijk} = \sum_{i=1}^{n_i} q_{ij}(t_0) e^{-(t_k - t_0) / \tau_{ij}} + \sum_{i=1}^{n_i} \left\{ \sum_{n=1}^{k} \left\{ (1 - e^{-\Delta t_n / \tau_{ij}})(f_{ij} I^{(n)}{}_i) e^{-(t_k - t_n) / \tau_{ij}} \right\} \right\} \tag{2.29}$$

The gains and time constants are again estimated using multivariable nonlinear regression process. The required objective function is the same as CMRP:

$$\min z = \sum_{k=1}^{n_t} \sum_{j=1}^{n_p} (q_{jk}{}^{obs} - q_{jk})^2 \tag{2.19}$$

where q_{jk} is the model-calculated total production rate of producer j at time step k as defined by either equation (2.25) or (2.27). This objective function is constrained by (2.30) shown below as well as some additional constraints.

$$f_{ij}, \tau_{ij} \geq 0 \tag{2.30}$$

In the CRMIP model, gain and time constants are constrained to be positive as shown in equation (2.30). If the CRMT result shows that gain for the whole field is equal or less than one, which means some injection may be lost, the constraint for each injector will be:

$$\sum_{j=1}^{n_p} f_{ij} \leq 1 \text{ for all } i \tag{2.21}$$

If the CRMT result shows that the gain for the whole field is strictly greater than one, this indicates aquifer or injection outside the reservoir may be present. In such a case, the constraint (2.21) for each injector will be ignored, which means the summation of gains $\sum_{j=1}^{n_p} f_{ij}$ of injector i can be less, equal or greater than 1.

The summary of the basic equations used in the capacitance resistance model is summarized in Table 2.1**.**

Table 2.1 Summary of equations in the CRM model

	Objective Function	Constraints: Equal	Constraints: Unequal	
CRMT	$\min z=\sum_{k=1}^{n_t}\left(\sum_{j=1}^{n_p} q_{jk}^{obs}-q_k\right)^2$	$q_k=q_{k-1}e^{-\Delta t/\tau}+(1-e^{-\Delta t/\tau})\, fI_k$ Or $q_k=q_{k-1}e^{-\Delta t/\tau}+(1-e^{-\Delta t/\tau})(fI_k-J\tau\frac{P_{wf}^{\,k}-P_{wf}^{\,k-1}}{\Delta t})$	$f>0$ $\tau>0$	
CRMP	$\min z=\sum_{k=1}^{n_t}\sum_{j=1}^{n_p}(q_{jk}^{obs}-q_{jk})^2$	$q_{jk}=q_{j(k-1)}e^{-\Delta t/\tau_j}+(1-e^{-\Delta t/\tau_j})(\sum_{i=1}^{n_i} f_{ij}I_{ik}-J_j\tau_j\frac{P_{wf,j}^{\,jk}-P_{wf,j}^{\,j(k-1)}}{\Delta t})$	Balanced	$\sum_{j=1}^{n_p} f_{ij}\le 1$ for all i $f_{ij},\tau_j\ge 0$
		Or $q_{jk}=q_{j(k-1)}e^{-\Delta t/\tau_j}+(1-e^{-\Delta t/\tau_j})\sum_{i=1}^{n_i} f_{ij}I_{ik}$	Unbalanced	$f_{ij},\tau_j\ge 0$
CRMIP	$\min z=\sum_{k=1}^{n_t}\sum_{j=1}^{n_p}(q_{jk}^{obs}-q_{jk})^2$	$q_{jk}=\sum_{i=1}^{n_i} q_{ijk}$ $=\sum_{i=1}^{n_i}\left[q_{ij(k-1)}e^{-\Delta t/\tau_{ij}}+(1-e^{-\Delta t/\tau_{ij}})(f_{ij}I_{ik}-J_{ij}\tau_{ij}\frac{P_{wf}^{\,jk}-P_{wf}^{\,j(k-1)}}{\Delta t})\right]$	Balanced	$\sum_{j=1}^{n_p} f_{ij}\le 1$ for all i $f_{ij},\tau_{ij}\ge 0$
		Or $q_{jk}=\sum_{i=1}^{n_i} q_{ijk}=\sum_{i=1}^{n_i}\left[q_{ij(k-1)}e^{-\Delta t/\tau_{ij}}+(1-e^{-\Delta t/\tau_{ij}})\, f_{ij}I_{ik}\right]$	Unbalanced	$f_{ij},\tau_{ij}\ge 0$

2.2 Data Quality Control Algorithm in the CRMP Model

Production data are the most abundant data in the field. However, they can often be of poor quality because of undocumented operational problems, changes in operating conditions, or even recording mistakes (Nobakht et al. 2009). If this poor quality or inconsistency is not recognized as such, it can be misinterpreted as a reservoir issue other than a data quality problem. Thus quality control of production data is a critical and necessary step that must precede any interpretation of production data.

The CRMP model is a quick reservoir model based on the continuity equation. It tells how the reservoir would respond under certain injection signals. The basic idea of data quality control is to apply the CRMP model to evaluate the correct production response under some injection scheme. By doing this, one can achieve the goal of correcting the problematic production data to realize data reconciliation. In the next section, the algorithm of the DQC in the CRMP model is discussed.

Often a complete measurement of bottom-hole pressure fluctuations over the production history is unavailable. However, for a mature water flood reservoir, the bottom-hole pressure does not change drastically and tends to be constant. In the next discussion, the CRMP working equation without the bottom-hole pressure term will be used. However, having bottom-hole pressure makes no difference to our method except to replace the equation with the version that has the bottom-hole pressure term. According to the CRMP model, the governing equation in the absence of bottom-hole pressure will be:

$$q_{jk} = q_{j(k-1)} e^{-\Delta t / \tau_j} + (1 - e^{-\Delta t / \tau_j}) \sum_{i=1}^{n_i} f_{ij} I_{ik} \tag{2.17}$$

The objective function and constraints are:

$$\min z = \sum_{k=1}^{n_t} \sum_{j=1}^{n_p} (q_{jk}{}^{obs} - q_{jk})^2 \tag{2.19}$$

$$\sum_{j=1}^{n_p} f_{ij} \le 1 \text{ for all } i \tag{2.21}$$

$$f_{ij}, \tau_j \ge 0 \tag{2.20}$$

An implicit prerequisite for this mathematical model, as well as other CRM models, to give the right model parameter is that the total production rate change of producer j, should be only caused by the reservoir itself instead of other reasons.

When only problematic production data are available, the following algorithm will be applied in the CRMP model to retrieve the corrected data:

$$\min z = \sum_{k=1}^{n_t} \sum_{j=1}^{n_p} \left(q_{jk}{}^{obs} - q_{jk}{}^{prob} \right)^2 \tag{2.31}$$

$$q_{jk}{}^{prob} = q_{j(k-1)}{}^{prob} e^{-\Delta t / \tau_j} + (1 - e^{-\Delta t / \tau_j}) \sum_{i=1}^{n_j} f_{ij} I_{ik} \tag{2.32}$$

$$\sum_{j=1}^{n_p} f_{ij} \le 1 \text{ for all } i \tag{2.21}$$

$$f_{ij}, \tau_j \ge 0 \tag{2.20}$$

where

$q_{jk}{}^{prob}$ is the calculated problematic production data of producer j at time step k

$q_{jk}{}^{obs}$ is the observed problematic production data of producer j at time step k

The reason why we call $q_{jk}{}^{prob}$ the “calculated-problematic” production data is because the observed data $q_{jk}{}^{obs}$ themselves are problematic. In other words, the objective function will produce model parameters to match the problematic data instead of the true production rates. And these model parameters will be used to recalculate the production rate, which is q_{jk}^{prob} in the equation. Problematic observed data give problematic gain and time constant and, as a result, problematic model-calculated production rate q_{jk}^{prob}.

The first application of the CRMP model using the original problematic data set gives two model parameters, gain and time constant which will be denoted by $f_{ij}^{(1)}$ and $\tau_j^{(1)}$. After identifying the problematic section of data (only some of the time periods in the data set are incorrect), the next step is to use the gain $f_{ij}^{(1)}$ and time constant $\tau_j^{(1)}$ together with the observed total production rate q_{jk}^{obs} to re-evaluate the production rate of the problematic time period using the following equations:

$$q_{jk}^{prob1} = q_{j(k-1)}^{obs} e^{-\Delta t / \tau_j^{(1)}} + (1 - e^{-\Delta t / \tau_j^{(1)}}) \sum_{i=1}^{n_j} f_{ij}^{(1)} I_{ik} \quad \text{For } k = k_{prob1} \qquad (2.33)$$

$$q_{jk}^{prob1} = q_{j(k-1)}^{prob1} e^{-\Delta t / \tau_j^{(1)}} + (1 - e^{-\Delta t / \tau_j^{(1)}}) \sum_{i=1}^{n_j} f_{ij}^{(1)} I_{ik} \quad \text{For } k_{porb1} < k < k_{prob-end} \qquad (2.34)$$

where k_{prob1} is the first problematic time step, and $k_{prob-end}$ is the last problematic time step. Equation (2.33) and (2.34) mean that the production at the very first problematic time step will be calculated from $q_{j(k-1)}^{obs}$ which is one step before the problematic time step. And this observed production $q_{j(k-1)}^{obs}$ is error free. The successive production rates after the very first problematic date will be calculated from last production rate calculated by equation (2.34) until the end of problematic time period.

q_{jk}^{prob1} calculated by equation (2.33) or (2.34) is still problematic because the gain and time constant used in equation (2.33) or (2.34) above are misleading considering the fact that they come from the process to fit the problematic data. However we call q_{jk}^{prob1} the corrected data in a sense that, although problematic, most portions of the production data are good data, which enables the CRM model to give model parameters that are close to the truth. q_{jk}^{prob1} will be used to place the problematic section of q_{jk}^{obs}. The section in $q_{jk}{}^{obs}$ without any problem will remain the same, while the error section will be replaced by $q_{jk}{}^{prob1}$. After this replacement, a new production rate $q_{jk}{}^{obs1}$ is obtained and will be used in the CRM model instead of the original production data $q_{jk}{}^{obs}$.

$$q_{jk}{}^{obs1} = \begin{cases} \text{Period of data without error for } k_{porb1} > k \; or \; k > k_{prob-end} \\ \\ \text{Period of problematic data replaced by } q_{jk}{}^{prob1} \text{ for } k_{porb1} < k < k_{prob-end} \end{cases}$$

where k_{prob1} is the first problematic date, and $k_{prob-end}$ is the last problematic period date.

The next step is to run the CRMP model again with $q_{jk}{}^{obs}$ replaced by $q_{jk}{}^{obs1}$ in the objective function (2.31) and constraints keep the same. After the second application of the CRMP model using the corrected data set $q_{jk}{}^{obs1}$, we obtain another gain $f_{ij}{}^{(2)}$ and time constant $\tau_j{}^{(2)}$, which will again be used to re-evaluate and then replace the problematic period. Again the section in $q_{jk}{}^{obs1}$ without problem remains the same while the error section will be replaced by $q_{jk}{}^{prob2}$ defined in equation (2.35) or (2.36).

$$q_{jk}{}^{prob2} = q_{j(k-1)}{}^{obs1} e^{-\Delta t / \tau_j{}^{(2)}} + (1 - e^{-\Delta t / \tau_j{}^{(2)}}) \sum_{i=1}^{n_j} f_{ij}{}^{(2)} I_{ik} \text{ For } k = k_{prob1} \tag{2.35}$$

$$q_{jk}{}^{prob2} = q_{j(k-1)}{}^{prob1} e^{-\Delta t / \tau_j{}^{(2)}} + (1 - e^{-\Delta t / \tau_j{}^{(2)}}) \sum_{i=1}^{n_j} f_{ij}{}^{(2)} I_{ik} \text{ For } k_{porb1} < k < k_{prob-end} \tag{2.36}$$

$$q_{jk}{}^{obs2} = \begin{cases} \text{Period of data without error for } k_{porb1} > k \; or \; k > k_{prob-end} \\ \\ \text{Period of problematic data replaced by } q_{jk}{}^{prob2} \text{ for } k_{porb1} < k < k_{prob-end} \end{cases}$$

where k_{prob1} is the first problematic date, and $k_{prob-end}$ is the last problematic date. After the second replacement, a new $q_{jk}{}^{obs2}$ is obtained and will be used in the CRMP model.

The next step is to run the CRMP model for the third time with $q_{jk}{}^{obs1}$ replaced by $q_{jk}{}^{obs2}$ in the objective function (2.31) and constraints keep the same. After this running of the CRMP model, we get $f_{ij}^{(3)}$ and $\tau_j^{(3)}$, and they will be used to re-calculated and then replace the problematic production rate once again.

Keep repeating this process until the gain and time constant converge. By "converge", we mean that gain and time constant does not change more than a given tolerance. The tolerance is defined as the absolute relative change of gain and time constant calculated from two consecutive application of the CRMP model. For example, after n times of running the CRM model, the gain $f_{ij}^{(n)}$ and time constant $\tau_j^{(n)}$ can be obtained. Comparing to the last results $f_{ij}^{(n-1)}$ and $\tau_j^{(n-1)}$, if they satisfy:

$$\left|\frac{f_{ij}^{(n)} - f_{ij}^{(n-1)}}{f_{ij}^{(n)}}\right| \le tolerance \tag{2.37}$$

and

$$\left|\frac{\tau_j^{(n)} - \tau_j^{(n-1)}}{\tau_j^{(n)}}\right| \le tolerance \tag{2.38}$$

then the data quality control(DQC) procedure will stop; otherwise the DQC procedure will keep going on until the absolute relative error is less than the tolerance. Figure 2.4 shows how time constant and gain change with the number of the DQC procedure and finally converge. The final corrected production rate will be calculated by the gain and time constants that converge.

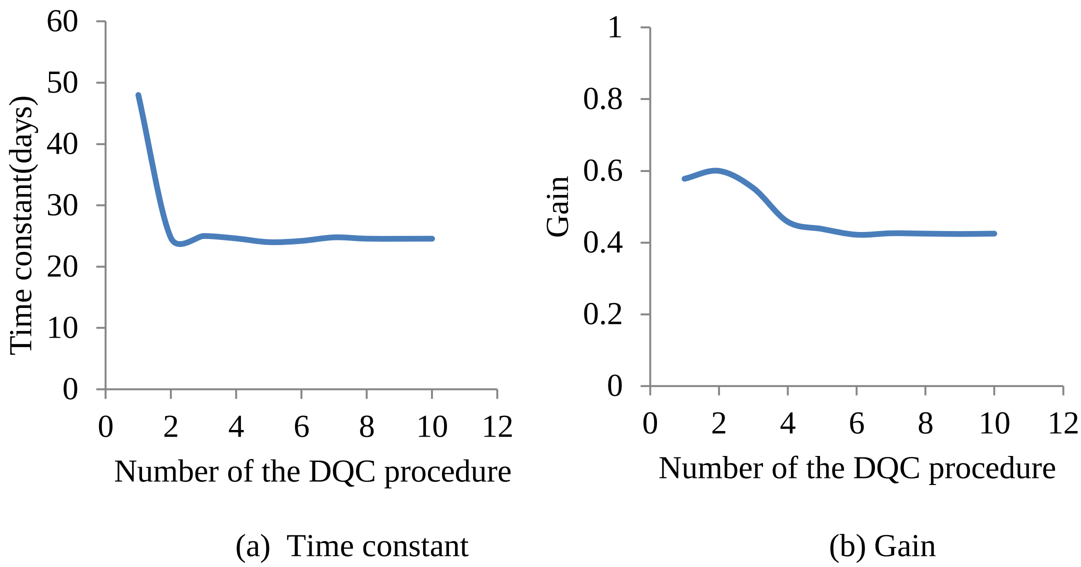


Figure 2.4 Time constant and gain change with the number of the DQC procedure

The CRMT and CRMIP follow the similar algorithms and equations in the data quality control process as the CRMP model. The summary of the data quality control process in the CRM model is summarized in Figure 2.5.

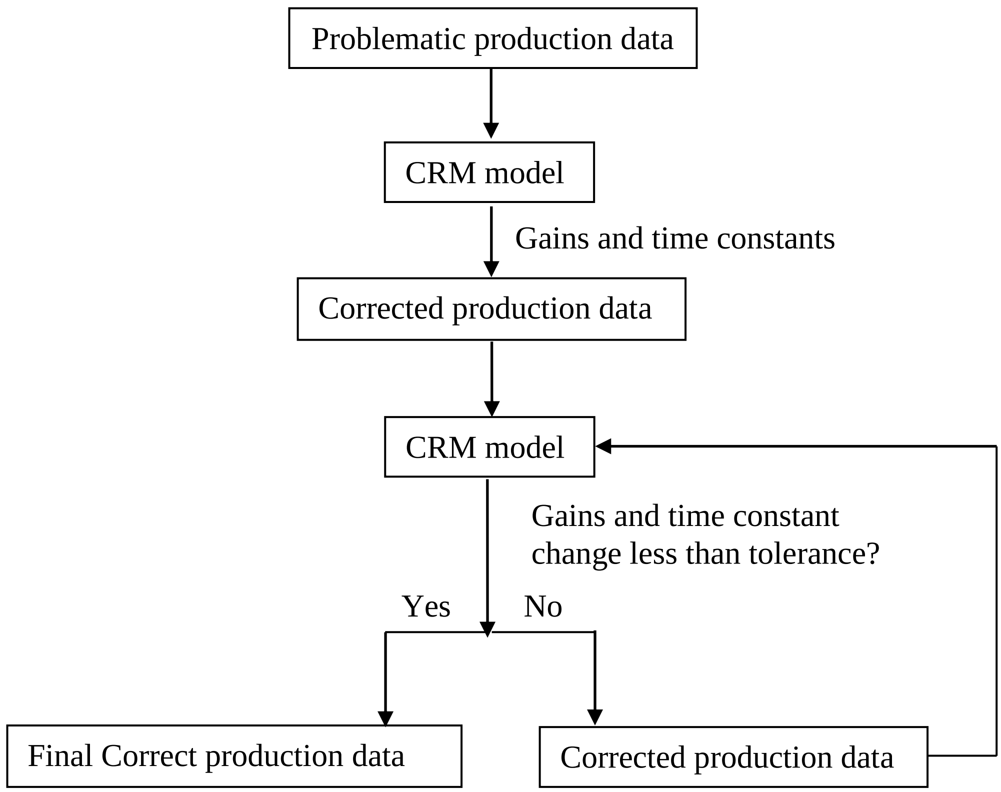


Figure 2.5 Summary of data quality control process in the CRM model

Chapter 3 Data Quality Control (DQC) in 5x4 Synthetic Field

In this chapter validation of the data quality control (DQC) process developed in Chapter 2 is presented. Different numerically simulated synthetic fields are used in the validation. The reason why we choose synthetic fields is because, unlike a real reservoir model where the geological structure and the reservoir properties must be specified, everything is known in a synthetic field. We are able to construct a synthetic reservoir with imposed geology and all other desired features. The production data from the synthetic field are free of error and are used as the base case to show the correct production response under some injection scheme.

In this and the next chapters we will perform sensitivity analysis in synthetic fields. The synthetic fields are of different sizes and features. There are two different synthetic fields used in the validation. They are:

1. 5x4 synthetic field with 5 injectors and 4 producers
2. 25x16 synthetic field with 25 injectors and 16 producers

This chapter will cover the 5x4 synthetic field. The 25x16 synthetic field can be found in Chapter 4.

3.1 Introduction to the Synthetic Fields

Synthetic fields are ideal heterogeneous reservoir models created in commercial simulators. In a synthetic field, one can easily generate production rate data by imposing injection rates on the reservoir. These production rate data are error-free since the production goes on smoothly without interruption, disturbance or recording mistakes. The error-free production rate data are referred to the base case.

In sensitivity analysis, we first corrupt or contaminate the production rate data from the synthetic field deliberately to create different scenarios of data quality problems. Then the DQC-corrected production data are obtained by applying the DQC procedure developed in the previous chapter to the corrupted production data. The effectiveness of this data reconciliation process is then evaluated by comparing production data after the quality control with what it is supposed to be before corruption (the base case). Another goal of sensitivity analysis is to find out what factors this quality control procedure is sensitive to, which in return, will better serve the purpose of data reconciliation.

3.2 Introduction to the 5x4 Synthetic Field

The 5x4 synthetic reservoir model is created in a commercial simulator (Eclipse E100 black oil simulator, 2005) that consists of 5 injectors and 4 producers, thus the name 5x4 synthetic field. Well locations of the 5x4 synthetic field are shown in Figure 3.1.

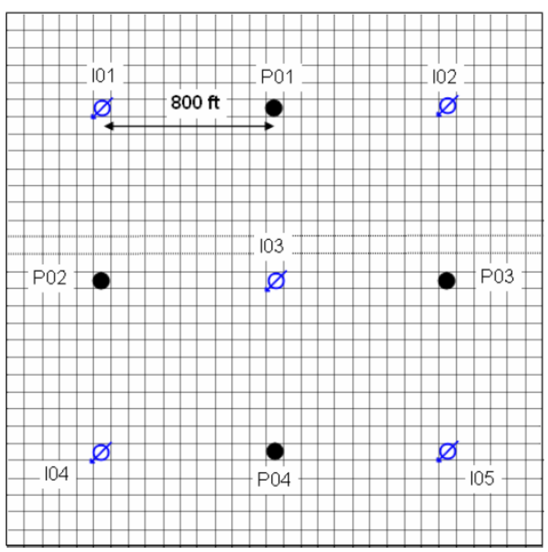


Figure 3.1: 5x4 synthetic field well locations

The 5x4 synfield has a 5-spot injection pattern with water and undersaturated oil. General information about dimensions of the synthetic field is listed in Table 3.1.

Table 3.1: 5x4 synthetic field dimensions

Parameters	Value
Number of grids	31x31x5
Grid sizes (ft)	80x80x12
Producer injector distance (ft)	800
Total reservoir volume (ft^3)	3.69×10^8
Total reservoir pore volume (ft^3)	6.642×10^7

The general reservoir properties of the 5x4 synthetic field are summarized in Table 3.2:

Table 3.2: 5x4 synthetic field reservoir properties

Property	Value
Porosity	0.18
Horizontal permeability (md)	40
Vertical permeability (md)	4
Oil compressibility (psi^{-1})	5×10^{-6}
Water compressibility (psi^{-1})	1×10^{-6}
Rock compressibility (psi^{-1})	1×10^{-6}
Initial reservoir pressure (psi)	1450
End-point oil-water mobility ratio	1

In this synfield, each producer is operating under the same constant bottom-hole pressure of 250 psi. The injection rates are fluctuating all the time. The monthly injection data (Figure 3.2) came from a real oil field and were scaled down to be proportional with the total injectivity of this synthetic reservoir model. The numerical simulation extends to 100 months, with one month for each time step.

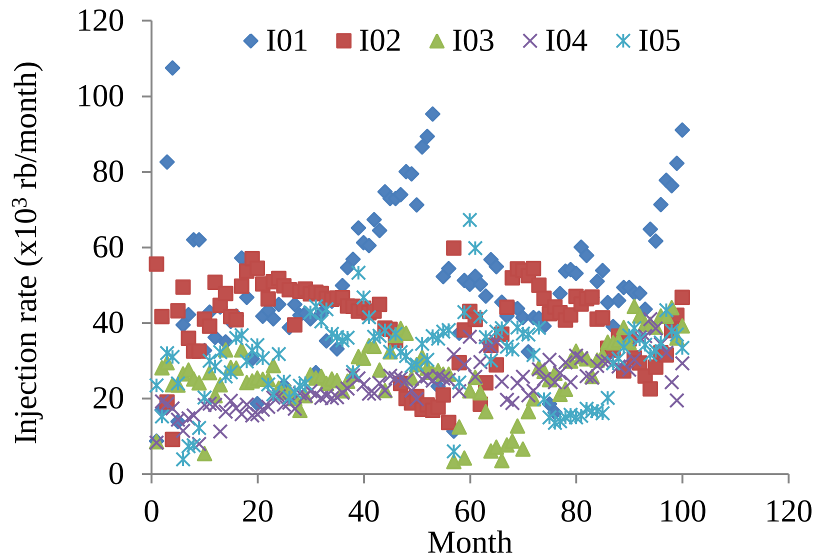


Figure 3.2: Injection rate data of the 5x4 synfield

Further, two high permeability streaks are imposed in the reservoir model. One streak is between well P01 and I01 with permeability equal to 1000md and the other is between well P04 and I03 with permeability equal to 500md as shown in Figure 3.3. This reservoir model will be referred to the "streak case" because of the high permeability streaks.

The existence of the two high permeability channels adds a heterogeneous feature to this simulated reservoir. After imposing injection rates shown in Figure 3.2 above to the streak case, a set of simulated production rates can be obtained. These production rates are the base case of the sensitivity analysis in this chapter.

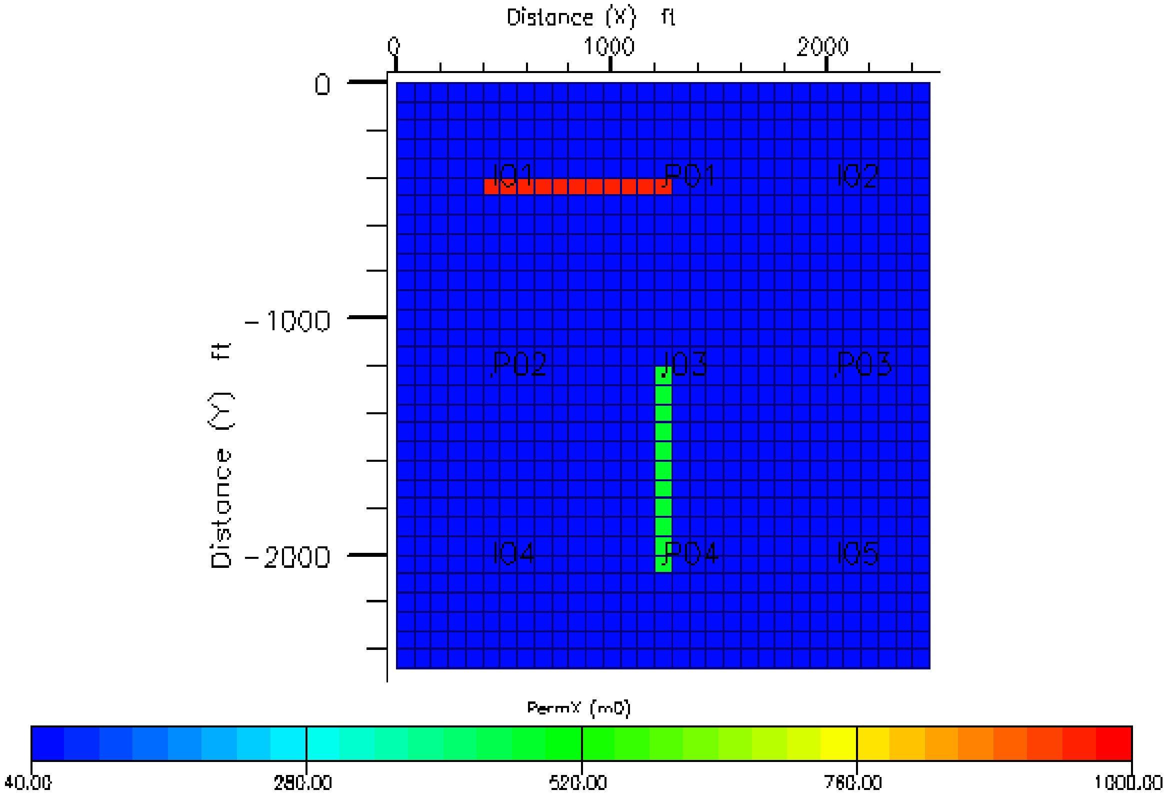


Figure 3.3: The 5x4 synfield –streak case

3.3 Sensitivity Analysis in the 5x4 Synthetic Field

3.3.1. Gains and Time Constant Sensitivity

In this section, attempts have been made to determine if gain and time constant values will affect the data quality control (DQC) procedure. From Chapter 2, one can conclude that, in general, a well with high connectivity usually has a small time constant while a well that is poorly connected with its neighbors tends to have a large time constant. In the next discussion, we are trying to find out if the DQC procedure is sensitive to the gain and time constant values.

To perform the sensitivity analysis, we first apply the error-free production data from the 5x4 synthetic field in the CRMP model. This step is designed to find out the connectivity and time constant corresponding to the uncontaminated production rates. The second step is to sabotage the production rate data deliberately and then apply the DQC procedure. Finally the relationship between gain (and time constant) and the effectiveness of the DQC procedure will be determined.

Applying the CRMP model to the 5x4 synthetic field first, the gain and time constant values are shown in Table 3.3.

Table 3.3: Gain and time constant for the 5x4 synthetic base case

	P01	P02	P03	P04
I01	0.89	0	0	0.11
I02	0.41	0.19	0.22	0.12
I03	0.15	0.14	0.10	0.61
I04	0.40	0	0	0.60
I05	0.23	0	0.17	0.54
Sum of gains	2.07	0.33	0.49	1.98
Producer time constant (days)	21.94	594.77	182.11	21.62

From the results of the CRMP model, wells P01 and P04 have high connectivity with neighbor injectors and their time constants are small, while well P02 and P03 are the opposite. These results are in a good agreement with the geology of the reservoir where P01 and P04 are located in the high permeability steaks while P02 and P04 are not. These 4 wells are chosen to show if the data quality control method is sensitive to gain and time constant values in the following sensitivity analysis.

To realize the goal of sensitivity analysis, one must make sure that the 4 producers in the 5x4 synfield are comparable. In the discussion of this section, all producers have been corrupted in the same way as well as in the same time periods to ensure the comparability. All producers have missing data (0 rate indicates missing data) from the time step 40 to 67 as seen in Figure 3.4.

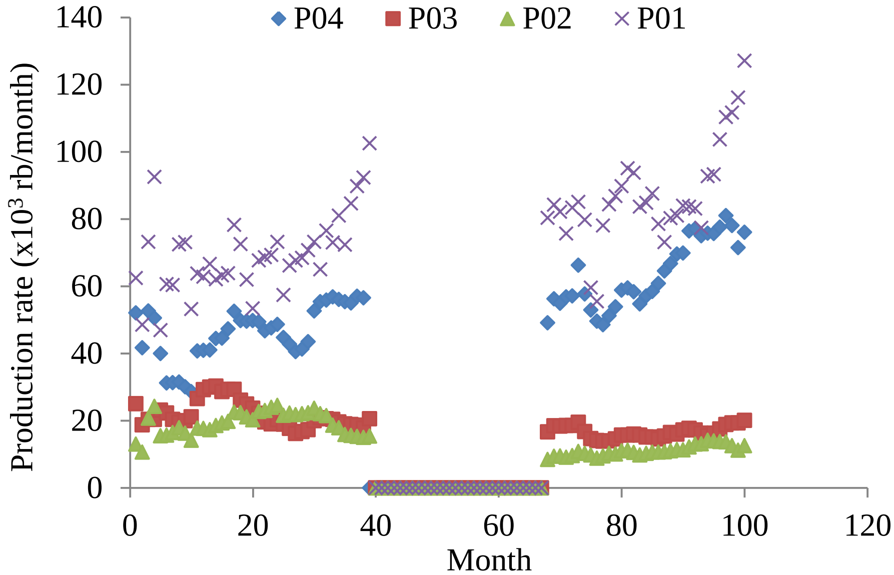


Figure 3.4: Producers with missing production rate data

However, for the purpose of analyzing gain and time constant values sensitivity, the production data can also be corrupted in other fashions (data averaging for example) as long as the corruption of production rates is consistent through all producers.

After applying the DQC procedure to all producers in the 5x4 synfield, the matching results of simulated production rates (the base case) versus the DQC-corrected production rate are shown in Figure 3.5.

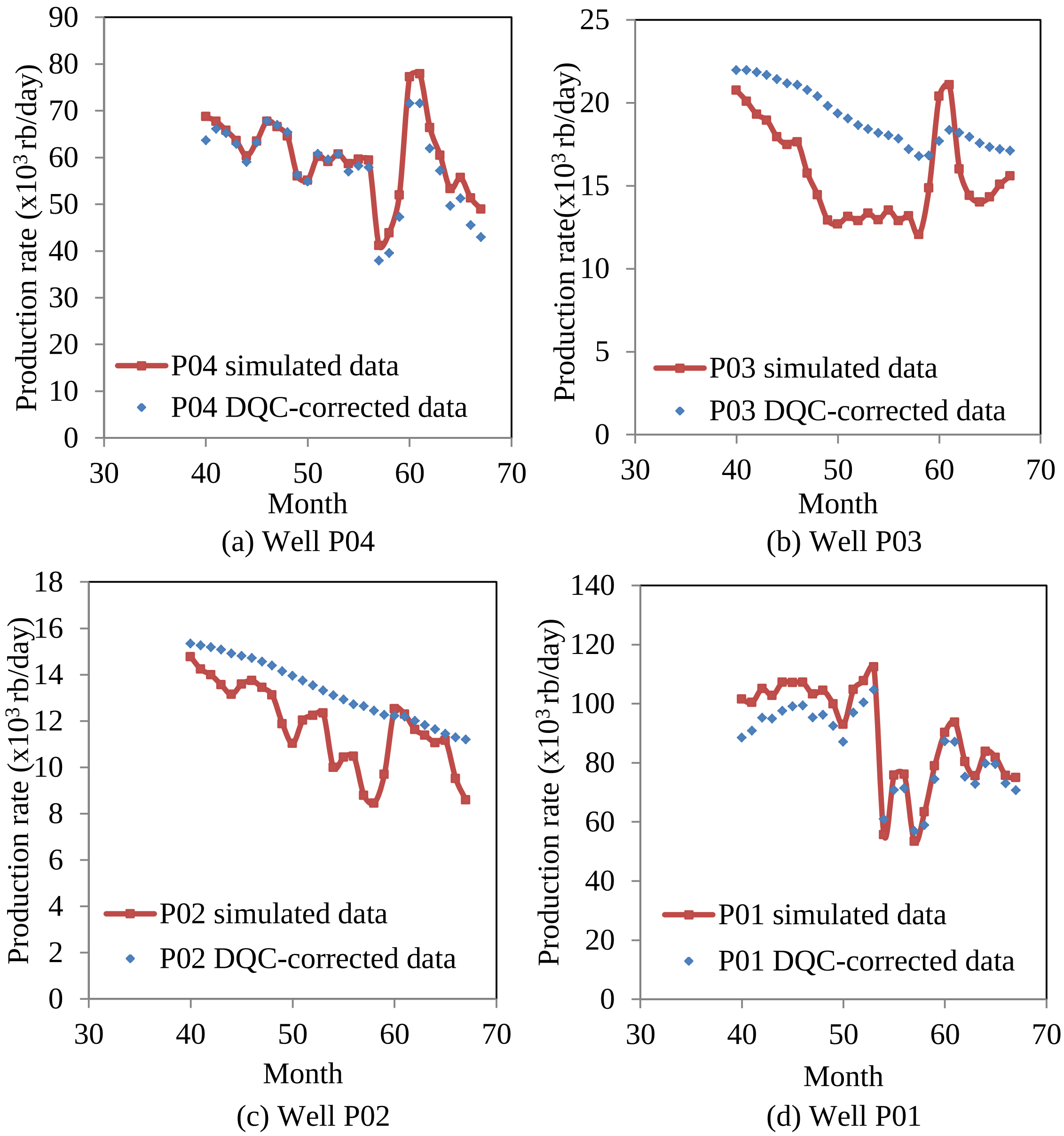


Figure 3.5: Simulated data vs. the DQC-corrected data in gain and time constant sensitivity analysis

Figure 3.5 reveals, in a straightforward way that, the DQC-corrected data of well P04 and P01 are closer to the simulated production data than those of well P02 and P03.

To quantify the efficiency of the DQC procedure, three different parameters are introduced.

The fitting quality of the DQC-corrected data to the simulated data (the base case) is described by the error between the model and the data. One may examine the absolute relative error by producer, defined as:

$$relerr_{jk} = \frac{\left| q_{jk}{}^{s} - q_{jk}{}^{c} \right|}{q_{jk}{}^{s}}$$

where

$q_{jk}{}^{s}$ is the simulated total production rate for producer j at time k

$q_{jk}{}^{c}$ is the DQC procedure corrected total production rate for producer j at time k

The absolute relative error by producer, $relerr_{jk}$ reveals the detailed fitting quality of each producer at each time step.

One may also examine the average absolute relative error by producer, defined as

$$err_{j} = \frac{1}{n_t} \sum_{k=1}^{n_t} \frac{\left| q^{s}{}_{jk} - q^{c}{}_{jk} \right|}{q^{s}{}_{jk}}$$

where n_t is the total number of the problematic periods.

The average absolute relative error by producer reflects the overall fitting quality through the fitting time period.

The third kind of relative error is the average relative error which is the summation of the average absolute relative errors of all producers divided by the number of producers.

$$avg\_rel = \frac{1}{n_j}\sum_j err_j$$

where n_j is the total number of producers.

The average relative error gives the general efficiency of the DQC procedure under some scenarios with data quality issues.

The relative errors in the production data missing scenario above are presented in Figure 3.6 and 3.7.

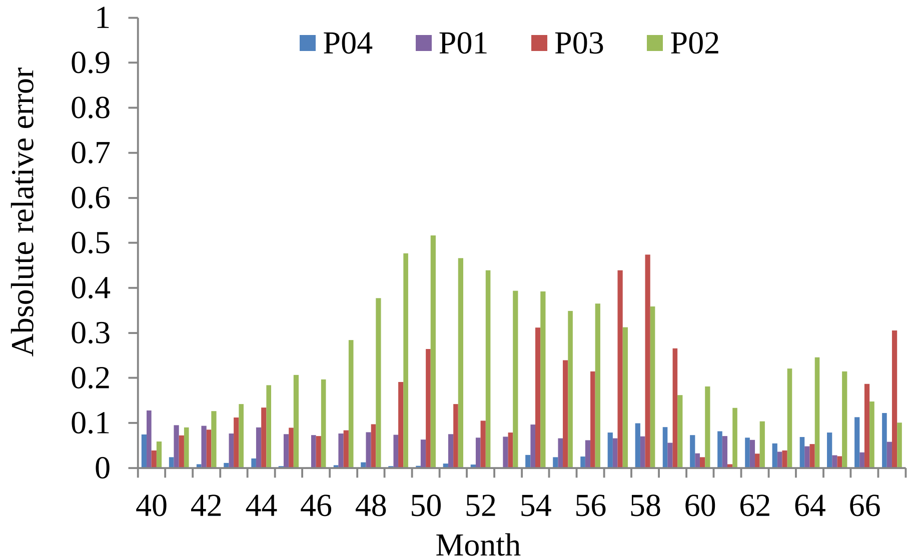


Figure 3.6: Absolute relative error using the DQC procedure in gain and time constant sensitivity analysis

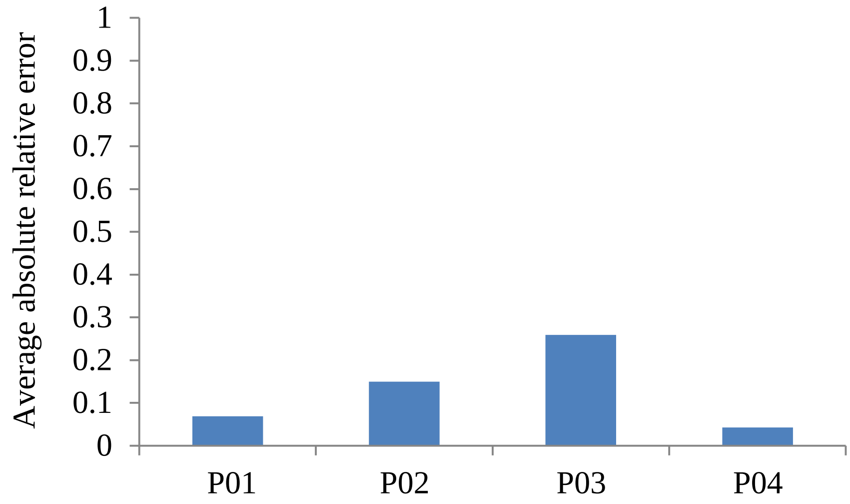


Figure 3.7: Average absolute relative error using the DQC procedure in gain and time constant sensitivity analysis

The average relative error figures show quantitatively that P04 and P01 have small relative errors (less than 6%) while well P02 and P03 have much larger ones, which indicates that the DQC procedure is more effective when applying on well P04 and P01.

The average absolute relative errors together with gain and time constant values are tabulated in Table 3.4.

Table 3.4: Gain and time constant with average absolute relative error in the 5x4 synfield

	Sum of gain	Time constant(days)	Date quality control (DQC)		
			Max relative error	Min relative error	Average relative error
P01	2.07	21.94	0.12	0.00	0.04
P02	0.33	594.77	0.47	0.01	0.15
P03	0.49	182.11	0.52	0.06	0.26
P04	1.98	21.62	0.13	0.03	0.07

Plots relating sum of gains and time constant with the average absolute relative errors are shown in Figure 3.8 and Figure 3.9.

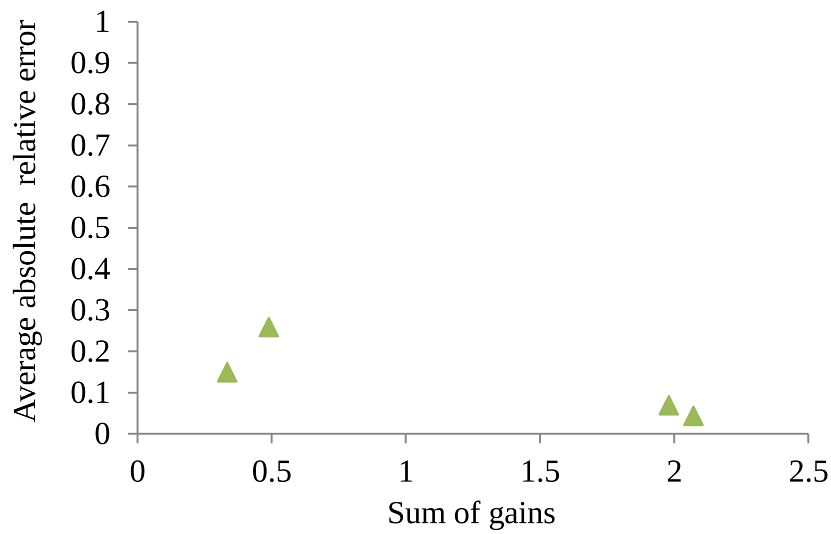


Figure 3.8: Change of average absolute relative errors with sum of gains in the 5x4 synfield

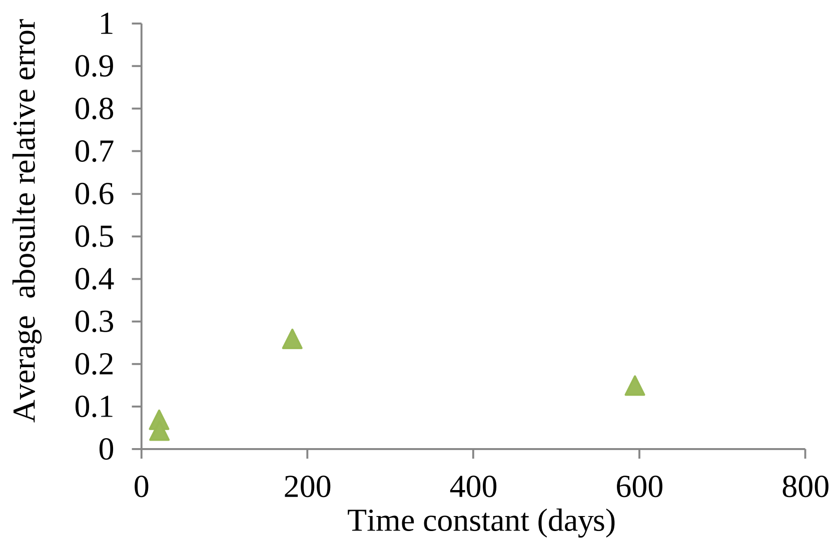


Figure 3.9: Change of average absolute relative errors with time constant in the 5x4 synfield

From the gain or time constant versus the average relative error plots, one can tell that the general trend is: the higher the connectivity (large sum of gain) the smaller the average absolute relative error, which indicates the efficiency of the DQC procedure (i.e. well P01 and P04). On the contrary, the large time constant comes with a large average absolute relative error, which is a sign of less effective DQC-correceted results (i.e. well P02 and P03).

However, there exists an exception that when sum of gain increases from 0.33 to 0.49 (time constant from 595 to 182 days), the average relative error, instead of decreasing, has increased. This abnormality may have been caused by the lack of sufficient samples since there are only 4 producers in this synthetic field. In Chapter 4 the results in a larger synfield with more producers will be presented to see if the conclusion we made in this small synfield still holds.

For summary, high connectivity (large sum of gains) comes with an efficient DQC-corrected production result. On the contrary, larger time constant often indicates less effective DQC-corrected production rates.

3.3.2. Number of Corrupted Periods

According to the algorithm of the DQC procedure, the correcting process of the problematic production rate is based on the fact that most portions of the production rates are good data, which enables the CRM model to give model parameters that are not far away from the truth. If the corruption is dominating the data set, the DQC procedure will be ineffective since the whole data set is misleading. Thus, it is important to decide a critical percentage of corruption in the DQC procedure. Using the critical percentage as a reference, if a data set is having more percentage of corruption than the critical one, then either the DQC procedure can't be used, or one must include more data to make it work.

To decide the critical percentage, we introduce two examples. Based on the sensitivity analysis about gain and time constant values above, well P04 is actively responding to the injection signal and thus is chosen to perform this task. By doing so, even if the DQC results on well P04 are not satisfying, it is only caused by the corruption percentage itself instead of the gain and time constant values.

In the following examples, well P04 is designed to lose different percentages of production data points. Then the goodness of fits of the DQC-corrected data to the simulated data is analyzed in the effort of deciding a critical percentage.

Example 1: Fixed missing data points, changing total number of data points

In this example, well P04 is missing prodction rate data from time step 39 to 59, which is 21 months in total. However the total number of data points used in the example is changing from 60 to 100 with 10 months increment once a time. This way, the percentage of missing data is changing accordingly from 21% to 35% (Table 3.5).

Table 3.5: Percentage of missing data points used in example 1

Case	Total data points	Missing data points	Percentage of missing data points (%)
1	60	21	35
2	70	21	30
3	80	21	26
4	90	21	23
5	100	21	21

Applying the DQC procedure to each scenario above, the average relative errors are shown in Figure 3.10 and Table 3.6.

Table 3.6: Summary of average relative errors in example 1 in number of corrupted data points sensitivity analysis

Percentage of missing data points (%)	Average relative error
21	0.02
23	0.03
26	0.10
30	0.10
35	0.08

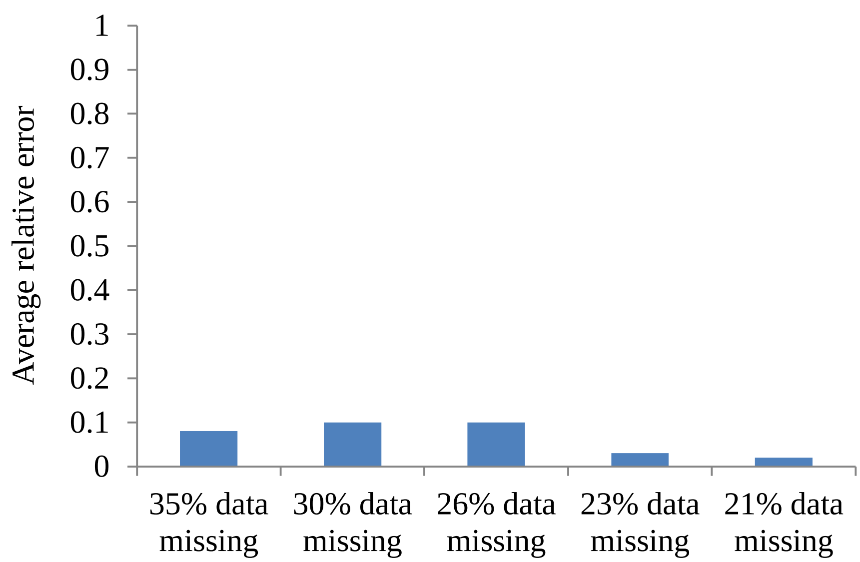


Figure 3.10: Average relative errors in example 1 using the DQC procedure in number of corrupted data points sensitivity analysis

As seen from Figure 3.10, the trend follows that the average relative error of fitting is increased by increasing the percentage of corrupted data. In other words, the efficiency of the DQC procedure gets worse when more production data get corrupted.

Notice that when there is only 21% of data missing (100 total months in Figure 3.10), the average raltive error is less than 2%. This small average raltive error indicates a good DQC-corrected data quality. However, there is one exception in Figure 3.10 that when only 60 total data points (which refers to 35% of data missing) are used, the average relative error decreases instead of increasing. This indicates that the trend that the DQC fit quality gets worse when increasing the corruption data percentage is not strictly satisfied. In other words, the average relative errors are increasing and at the same time fluctuating if the percentage of corruption increases.

Example 2: Fixed total number of data points, changing missing data points

Opposite to example 1, in example 2 different number of data points are missing while the total number of data points used remains the same (Table 3.7). Similarly, applying the DQC procedure, the average relative errors of the DQC-correced data vs. the simulated data under different corruption percentages are shown in Figure 3.11 and Table 3.8.

Table 3.7: Percentage of missing data points used in example 2

Case	Total data points	Missing data points	Percentage of missing data points (%)
1	100	11	11
2	100	21	21
3	100	31	31
4	100	41	41
5	100	51	51

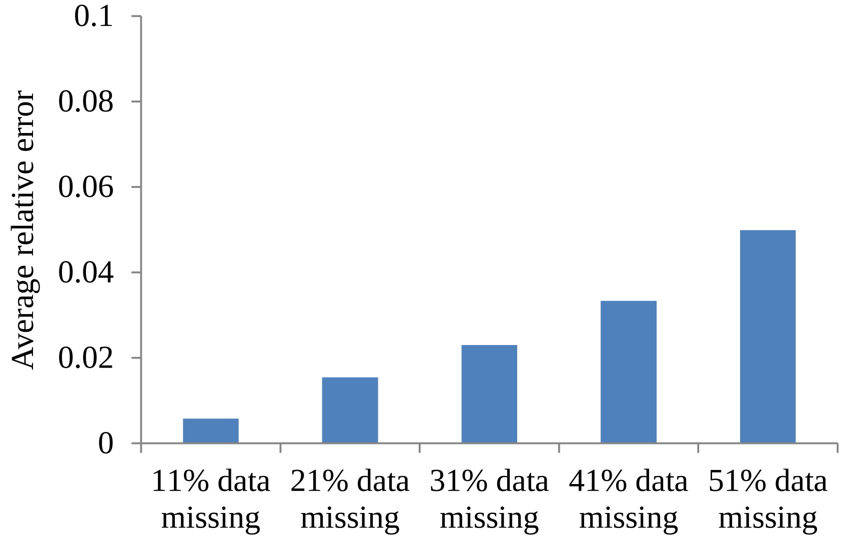


Figure 3.11: Average relative errors in example 2 using the DQC procedure in number of corrupted data points sensitivity analysis

Table 3.8: Summary of average relative errors in example 2 in number of corrupted data points sensitivity analysis

Percentage of missing data points (%)	Average relative error
11	0.01
21	0.02
31	0.06
41	0.09
51	0.11

Figure 3.11 shows a clear trend that when increasing the number of missing months, the average relative errors increase. In both example 1 and example 2, 21% of data points missing would refer to the same average relative error of 0.02 in Table 3.6 and Table 3.8. The change of average relative errors in Figure 3.11 reveals that: below 20% of data points missing, the average relative errors are all very small (less than 2%), and

that for different data missing percentage the average relative errors are similar to each other. However, above 20% of data points missing, it shows a sudden increase in the average relative errors, which futher makes the 20% a significant percentage for the data missing scenario.

Based on the analysis above, we propose to use the 80/20 rule in the DQC procedure. The 80/20 rule originated by Vilfredo Pareto who was an influential Italian economist. He attempted to turn economics into an exact science describable by laws. Through the empirical observation in his garden, he discovered that 80% of peas were produced by 20% of peapods. Because Italy's land was owned by 20% of the population, Pareto's application of the 80/20 rule grew from there. Briefly, the 80/20 rule means, for many events, roughly 80% of the effects come from 20% of the causes. For our data quality control, the 80/20 rule is a rule of thumb. However it has a new meaning in the new context. It means that 20% is a critical percentage of corrupted data. In other words, only 20% or less bad data is acceptable in the DQC process. The other 80% must be good data for the method to work.

3.3.3. Different Modes of Corruption

Production data can be corrupted in different modes. In this section, it will be discussed if different fashions of corruption will affect the results of a DQC process. Although the reason of data contamination or corruption hasn't been studied thoroughly, it is still helpful to categorize different kinds of data quality problems. Briefly, we summarize data quality problems in 3 main categories as following:

1. Production rate missing
2. Production rate averaging

3. Production rate sudden change

However, these problems can appear anywhere in a data set. The possible combinations of data quality problems together with the positions they may appear in the data sets will be discussed in the following contents. However, we often don't get to choose how and where the data are corrupted, the sensivity analysis in this chapter would give an overall confidence limit to expect when it comes to different type of data quality problems.

We will pick well P04 as the candidate to perform this analysis. By deliberately corrupting the simulated production data, one can create sets of problematic data with different corruption modes. The DQC-corrected data are obtained by applying the DQC process to these problematic data sets. Then comparisons are made between the DQC-corrected data and the simulated data (the base case) to reveal how the DQC is sensitive to the corruption modes.

Case A: Production rate missing

There are 3 different missing data scenarios:

(1) Missing data points in the front of the data set: in this scenario, producer P04 is missing data points (0 rate indicates missing data) from time step 1 to 20.

(2) Missing data points in the middle of the data set: P04 is missing data points from time step 39 to 59.

(3) Missing data points at the end of the data set: P04 is missing data points from time step 80 to 100.

Applying the data quality control process, the relative errors of well P04 under each situation are shown in Figure 3.12 and 3.13:

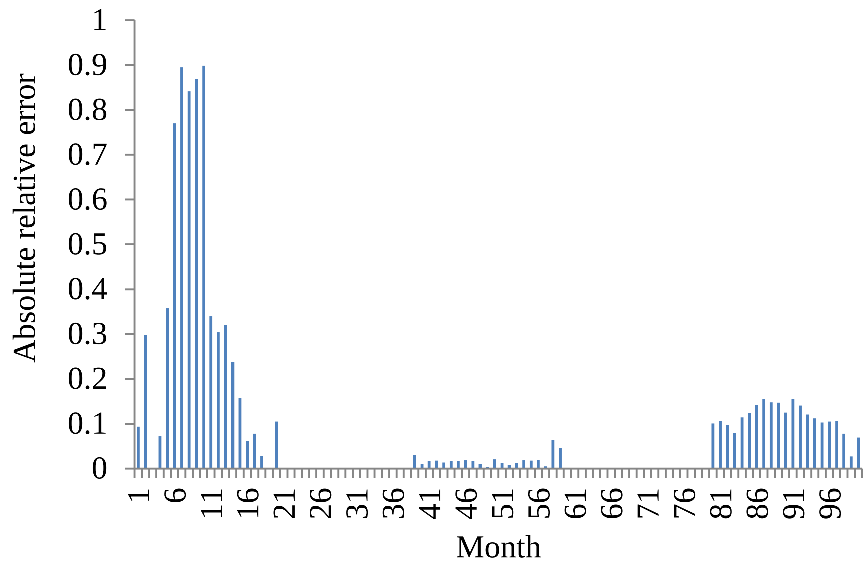


Figure 3.12: The absolute relative error caused by production rate missing in different part of the data set

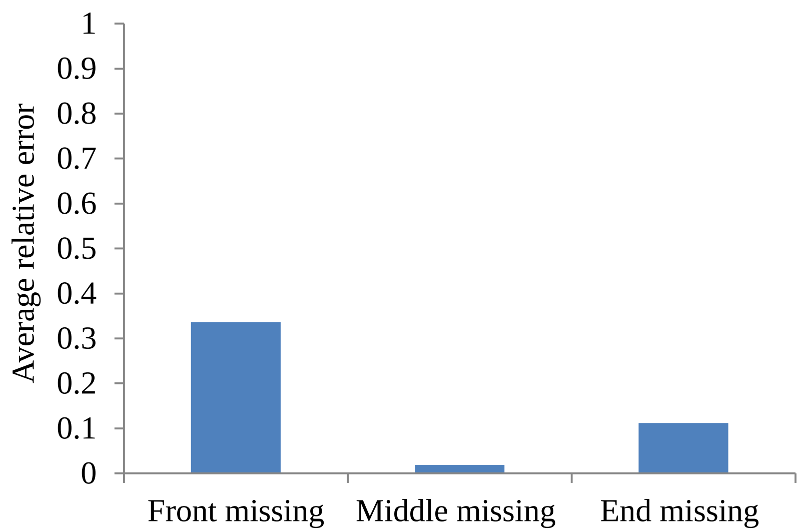


Figure 3.13: The average relative error caused by production rate missing in different part of the data set

Figure 3.12 and Figure 3.13 show the relative errors of fitting under different data missing scenarios. The results show that when data is missing in the middle of the data set, the average relative error is only 1.8% while if data is missing from the front part of the data set, the average relative error is about 34% which indicates overall less satisfactory DQC-retrieved data quality.

Figure 3.14 shows how the DQC-corrected production rates match the original simulated production rates. We put the matching results of the simulated rates vs. the DQC corrected rate under different production data missing scenarios in the same plot for well P04to show , straightforwardly, the difference in fits quality.

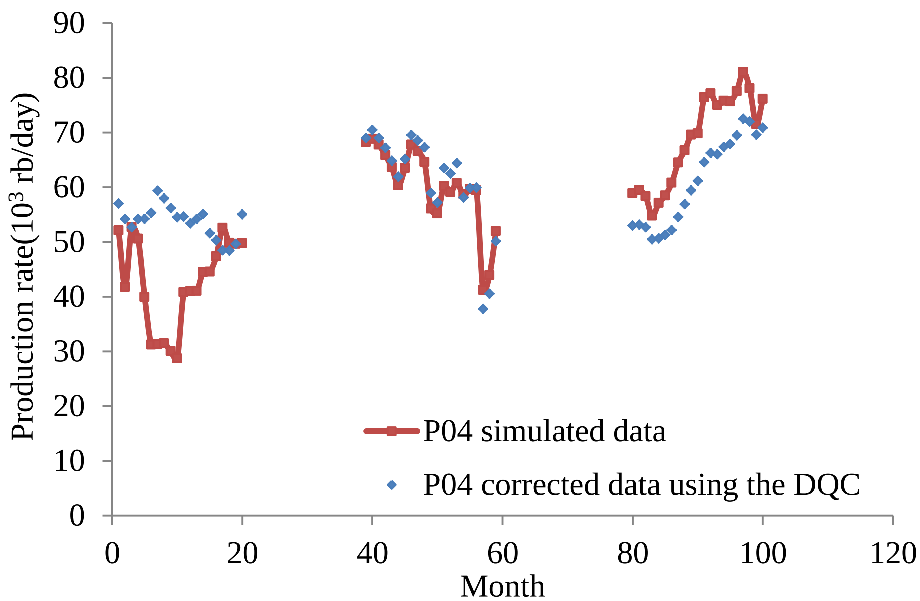


Figure 3.14: Simulated data vs. the DQC-corrected data of production rate data missing from different part of the data set of well P04

Figure 3.14 shows, in a straightforward way, that when data are missing in the middle of the data set, the DQC procedure will be able to retrieve data almost perfectly while when the data is missing from the end of the data set, the DQC results are less satisfying. However, under the front data missing scenario, while the original rates fluctuate around 55000 rb/day, the DQC-retrieved rates have suddenly decreased to less than 30000 rb/day, which reveals that the DQC process has failed to capture the original data in this scenario.

Case B: Sudden change of production rate

There are 3 different scenarios:

(1) Sudden rate change in the front of the data set: in this case, instead of normal fluctuations of production rates, P04 has a jump in the production rates (Figure 3.1 5), either increase or decrease from time step 1 to 21.

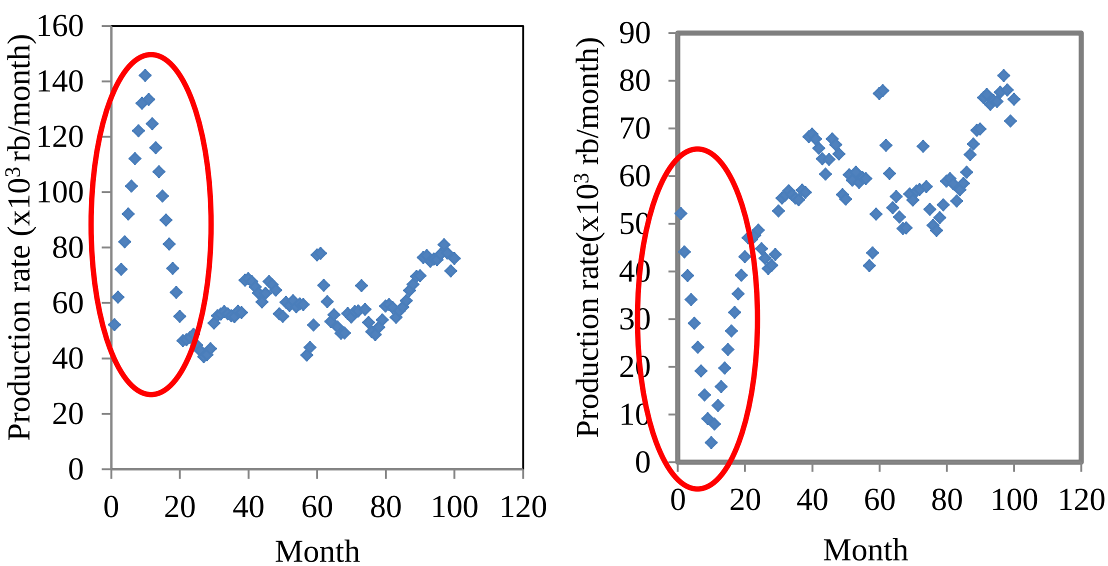


Figure 3.15: Production rate sudden linear change in the front of the data set

Applying the DQC process, the absolute relative errors of well P04 at each time step are shown in Figure 3.16.

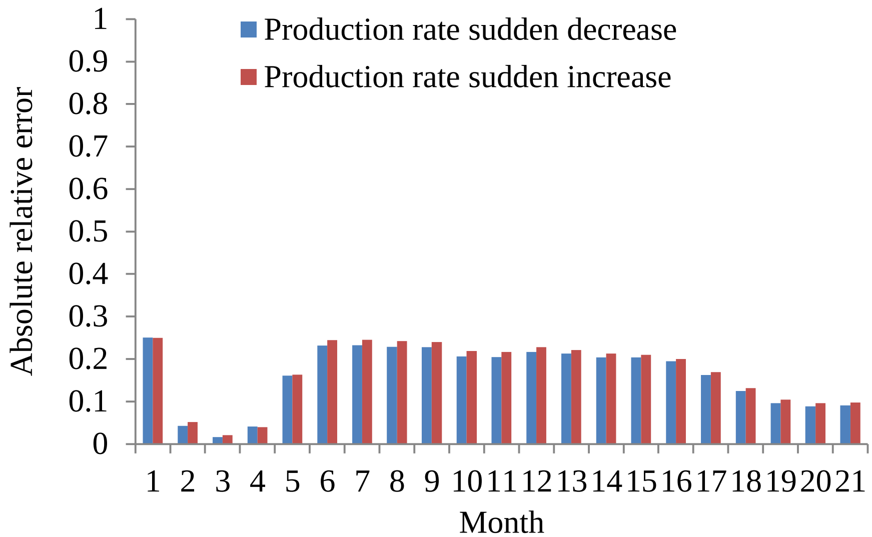


Figure 3.16: The absolute relative error of the DQC procedure caused by production rate sudden linear change in the front of the data set

Figure 3.16 shows that both sudden increase and decrease scenarios give similar absolute relative errors. However, as shown in Figure 3.15, the production rates are corrupted in a linear fashion, which is not the case when it comes to the real oil field production data. The real production rate data are often fluctuating. The mathematically random-looking characteristic of real production rates has invoked the interest to see if the data have been corrupted in a random fashion, how the DQC procedure is going to work.

For the purpose of testing the DQC procedure under a realistic scenario, well P04 is having a sudden increasing or decreasing of random production rates (Figure 3.17). All the random rates are generated using the “RAND” function in Excel 2007. In Excel, the “RAND” function returns a random number that is greater than or equal to 0 and less than 1. We take a set of random number generated by “RAND”, multiply them with different large numbers (100, 1000 or even 10000) to be proportional to the production rates in the synthetic field as seen in Figure 3.17. Applying the DQC process to this random rate change scenario, the absolute relative errors are shown in Figure 3.18.

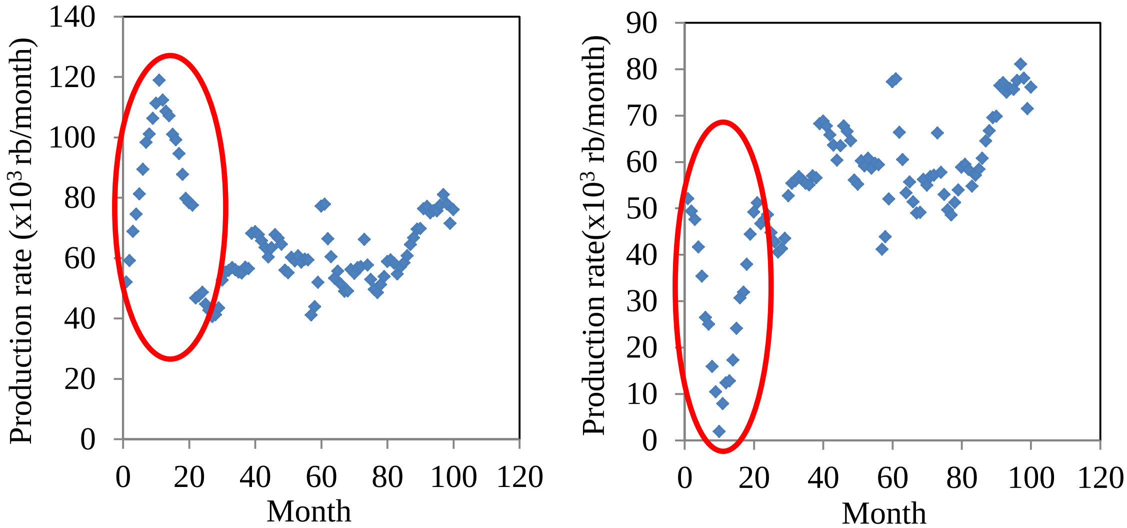


Figure 3.17: Production rate sudden random change in the front of the data set

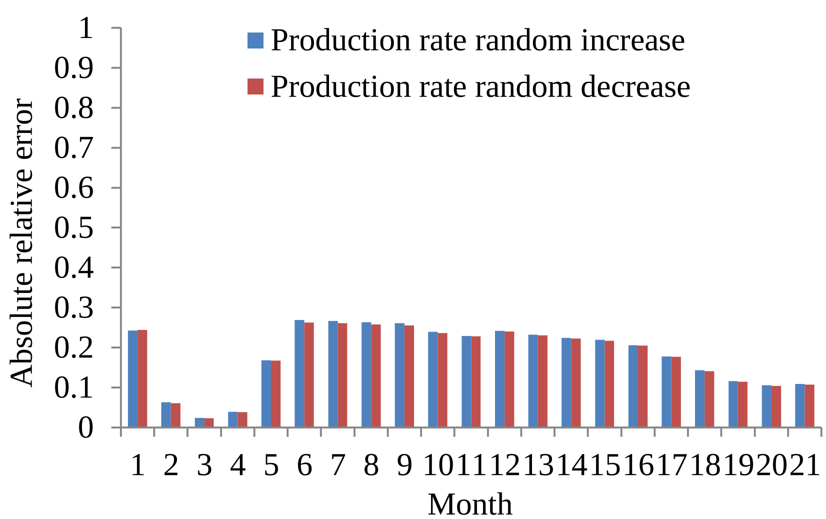


Figure 3.18: The absolute relative error of the DQC procedure caused by production rate sudden linear change in the front of the data set

Figure 3.18 looks similar to Figure 3.16 where the DQC-corrected results are again consistent under scenarios of increase and decrease.

So far, we have tested the possibility of the sudden data change that happens in the front part of the data set. Figure 3.19 summarizes the absolute relative errors of all scenarios. In Figure 3.19, the relative errors are not small. Most of the absolute relative errors are over 20%. Although consistent, the overall DQC-corrected data quality is not satisfying when the sudden rate change happens in the front of the data set.

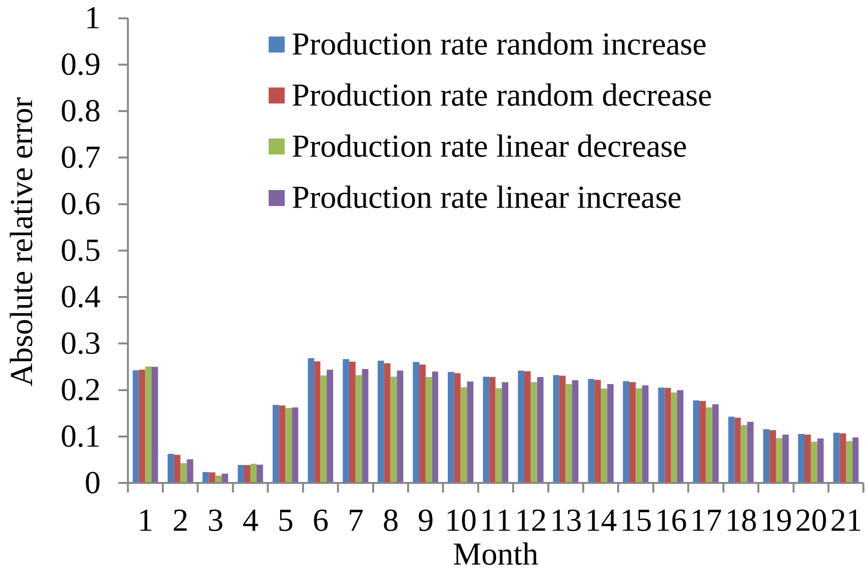


Figure 3.19: Summary of absolute relative error of the DQC procedure caused by sudden production rate change in the front of data set

The average absolute relative errors of each scenario are summarized in Figure 3.20 and Table 3.9. The average relative errors are similar to each other. These results reveal that if the sudden change of production rates happens in the front of a data set, one should expect the DQC procedure to give average absolute relative errors from 16% to 20%. Similarly, we also test the sudden data change in the middle and end of the data set.

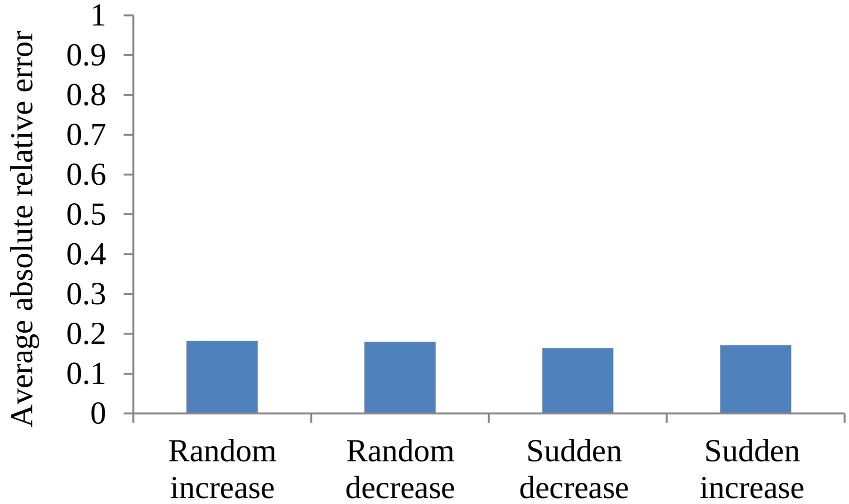


Figure 3.20: The average relative error of the DQC procedure caused by sudden production rate change in the front of data set

Table 3.9 Average absolute relative error of the DQC procedure caused by sudden production rate change in the front of data set

	Average relative error
Front random increase	0.183
Front random decrease	0.180
Front linear decrease	0.164
Front linear increase	0.172

(2) Sudden rate change in the middle of the data set: in this case, P04 has either an increase or a decrease in the rates in the middle of the data set. Applying the DQC process, the absolute relative errors are shown in Figure 3.21. Figure 3.22 and

Table 3.10 show the average relative errors.

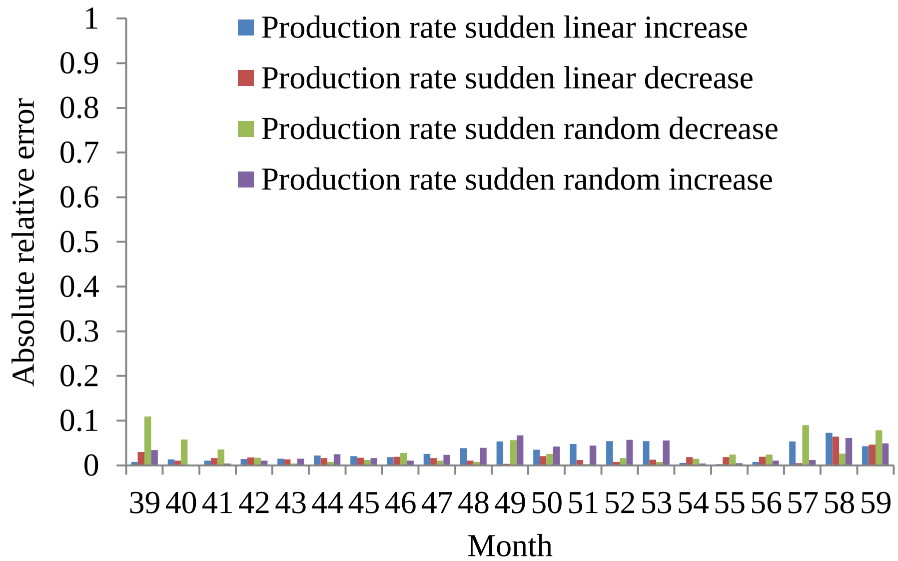


Figure 3.21: The absolute relative error of the DQC procedure caused by sudden production rate change in the middle of data set

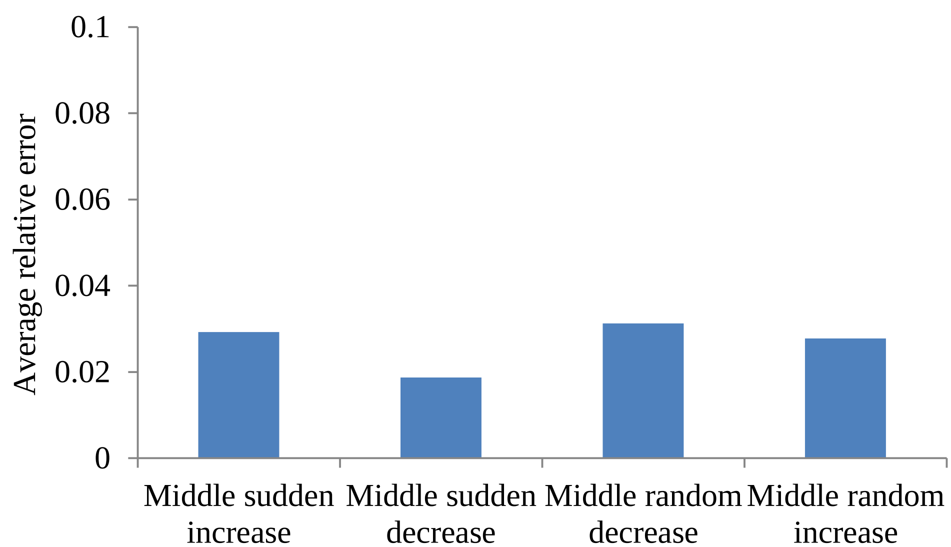


Figure 3.22: The average relative error of the DQC procedure caused by sudden production rate change in the middle of data set

Table 3.10 Average relative error of the DQC procedure caused by sudden production rate change in the middle of data set

	Average relative error
Middle sudden increase	0.029255
Middle sudden decrease	0.018742
Middle random decrease	0.031247
Middle random increase	0.027806

Figure 3.21 shows that the absolute relative errors of the DQC procedure are quite different for different scenarios. However, notice that in Figure 3.22 and Table 3.10, the average relative errors are very small, all around 3%, which indicates that if the sudden change of data happens in the middle of the data set, one may expect the DQC procedure to give average relative errors to be around 3%.

(3) Sudden rate change at the end of the data set: in this case, instead of normal fluctuations of production rates, well P04 has a sudden rates change, either an increase or a decrease at the end of the data set. Applying the DQC procedure, the absolute relative errors are shown in Figure 3.23. Figure 3.24 and Table 3.11 show the average relative errors.

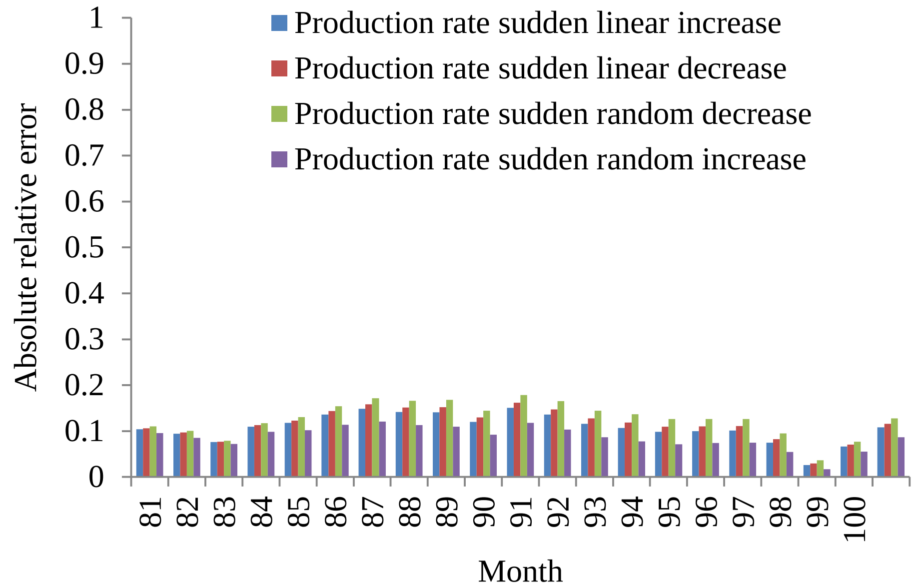


Figure 3.23: The absolute relative error of the DQC procedure caused by sudden production rate change at the end of data set

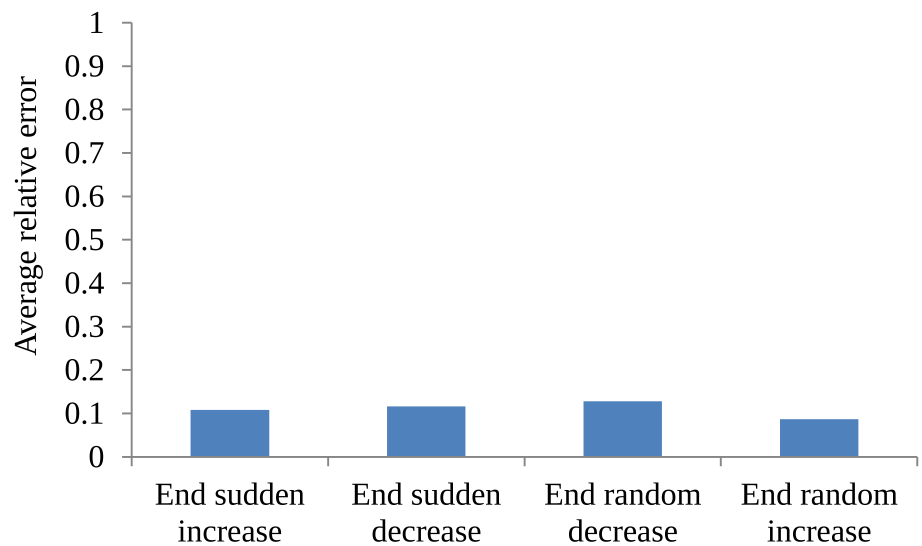


Figure 3.24: The average relative error of the DQC procedure caused by sudden production rate change at the end of data set

Table 3.11 Average relative error of the DQC procedure caused by sudden production rate change at the end of data set

	Average relative error
Back sudden increase	0.10836705
Back sudden decrease	0.116070213
Back random decrease	0.127753207
Back random increase	0.086765181

The average absolute relative errors of the DQC procedure under the sudden rates change at the end of the data set show little difference when it comes to different scenarios in Figure 3.24. From Table 3.11, the average absolute relative errors are small, all around 10%.

To summarize the sensitivity analysis of the production rates sudden change problems, Figure 3.25 below shows the overall fitting quality of DQC-corrected data one would expect when sudden rates change happens to different part in the data set.

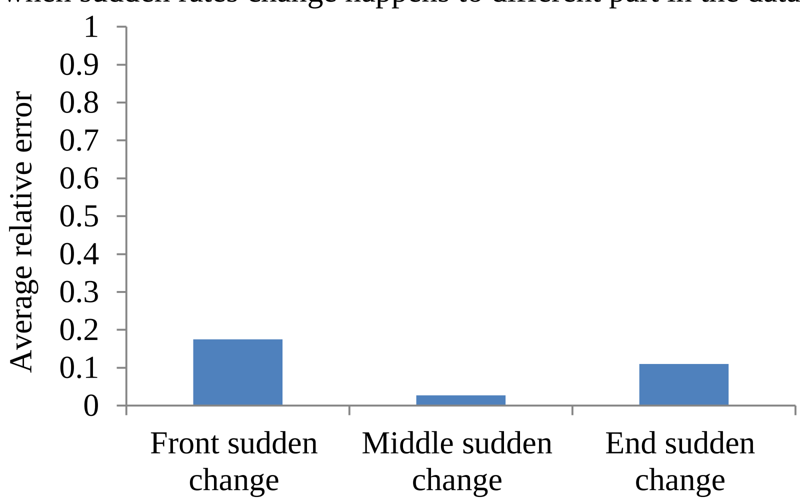


Figure 3.25: The average relative error of the DQC procedure caused by sudden production rate change in different parts of data set

Figure 3.25 reveals that, if possible, it is better to put the problematic data in the middle of the data set to get the best DQC-corrected results. However, if that is not going to happen, one may refer to Figure 3.25 to expect how much difference will occur when the problematic rates is in the front or end of the data set.

Case C: Production rates averaging

Flat (constant) production rates usually indicate infrequent rate measurement. This means that production rate is recorded per week or month; however it is reported as a constant value until another reading is taken (Nobakht et al. 2009).. To create this rate averaging scenario, instead of fluctuations of production rates, producer P04 has constant production rates of 50000 rb/month in the front (from time step 1 to 20), middle (from time step 39 to 59) and end (from time step 81 to 100) of the data set. Applying the DQC procedure, the absolute relative errors of well P04 at each time step are shown in Figure 3.26. The average absolute relative errors are shown in Figure 3.27.

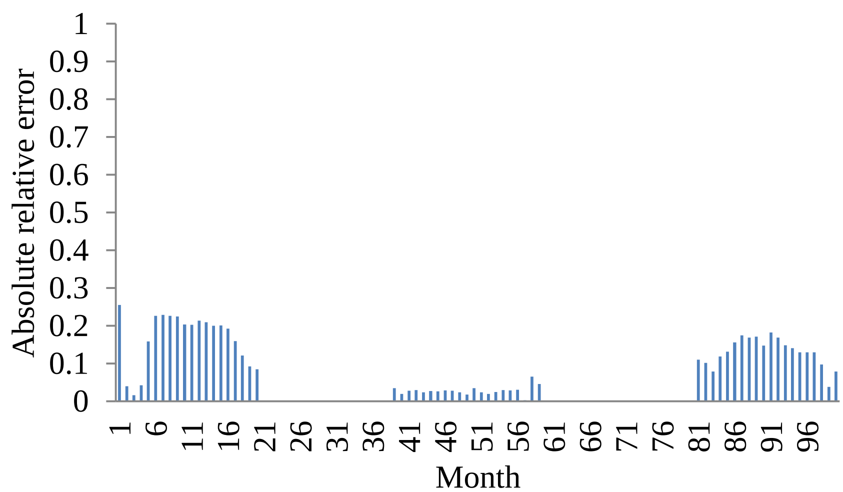


Figure 3.26: The absolute relative error of the DQC procedure caused by production ratea averaging in different parts in the data set

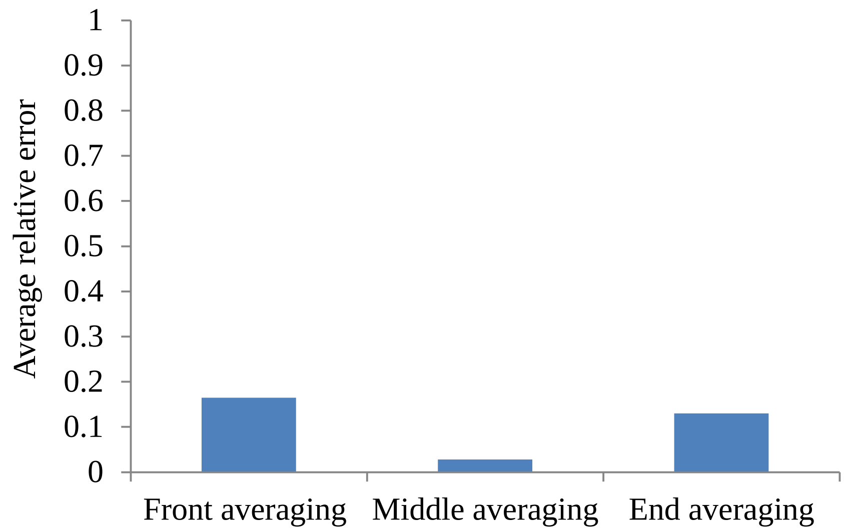


Figure 3.27: The average relative of the DQC procedure caused by production rate averaging in different parts in the data set

Figure 3.26 shows clearly that the middle part of the data set is the position where the DQC procedure works the best. In the front part of the data set, the DQC procedure gives worst results under the data averaging scenario. Figure 3.27 shows the average absolute relative errors for the middle rates averaging scenario are as low as 5%.

Comparing with Figure 3.25, which is the average absolute relative errors of sudden rates change, with Figure 3.27 above, they give similar average absolute relative errors despite of different modes of corruption.

Here we only show the results when the averaging rates are 50000 rb/month. Different values of the averaging rate are used to see if the DQC procedure is sensitive to the averaging values. The results show that the DQC process is sensitive to the position where the averaging rates appear rather than the averaging value itself.

3.3.4 Summary and Conclusion

In this chapter, we perform sensitivity analysis in the 5x4 synthetic field. Different data quality problems are presented to test the sensitivity of the DQC procedure. Attempts have been made to reach some conclusions as following:

1. Gain and time constant

In general, high connectivity (large sum of gain) often indicates good quality of DQC-corrected production rates. On the contrary, large time constant often comes with less satisfying DQC-corrected results.

2. Number of corrupted data points

For the purpose of data quality control, the 80/20 rule is used as a rule of thumb. It means that 20% is a critical maximum percentage of corrupted data. In other words, only 20% bad effect is acceptable in the DQC process. The other 80% has to be good data in order for the DQC procedure to work.

3. Different modes of corruption

The main goal to conduct sensivity analysis to different modes of corruption is to decide where to put the problemetic data to get the best DQC results and under which mode of corruption the DQC procedure works the best. However, when we don't get to choose where and how the data are corrupted, the sensivity analysis in this chapter would give an overall confidence limit that one would expect when it comes to different modes of corruption.

(1) Summarize by the position where the problematic data appears in the data set: Table 3.12, Table 3.13 and Table 3.14 summarize the average absolute relative errors of each mode of data corruption in the front, middle and end of

the data set.

Table 3.12 Average relative error of the DQC procedure caused by different data quality issues in the front of the data set

Date quality problems in the front	Average relative error
Missing	0.34
Averaging	0.16
Random increase	0.18
Linear increase	0.17
Linear decrease	0.16
Random decrease	0.18

Table 3.13 Average relative error of the DQC procedure caused by different data quality issues in the middle of the data set

Date quality problems in the middle	Average relative error
Missing	0.019
Averaging	0.028
Linear decrease	0.019
Random decrease	0.031
Random increase	0.028
Linear increase	0.029

Table 3.14 Average relative error of the DQC procedure caused by different data quality issues at the end of the data set

Date quality problems at the end	Average relative error
Missing	0.112
Averaging	0.130
Random decrease	0.128
Linear decrease	0.116
Linear increase	0.108
Random increase	0.087

Table 3.15 shows the average relative errors at different positions where the problematic production rates appear.

Table 3.15 Average relative error of the DQC procedure caused by different data quality issues at different parts of the data set

Position in the data set	Average relative error
Front	0.20
Middle	0.03
End	0.11

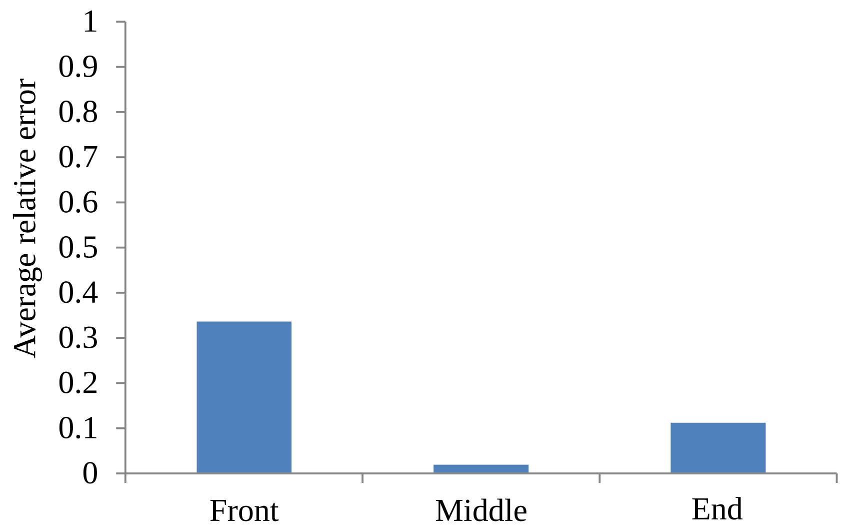


Figure 3.28 Average relative error of the DQC procedure caused by different data quality issues at different parts of the data set

From Table 3.15 and Figure 3.28, one can tell that the middle part of the data set is having the best quality of the DQC-correced results. Thus it is better to put the problematic data in the middle of the data set to get the best DQC-retrieved results and avoid putting them in the front of the data set where the DQC-correced results are less satisfying. However, if that is not going to happen, one could refer to Table 3.15 to know how much difference will occur after applying the DQC procedure when the problematic data is in the front or at the end of the data set.

(2) Summarize by the different modes of corruption : summrize the average relative errors with respect to different fashions of corruption in Figure 3.31 and Table 3.16 below:

Table 3.16 Average relative error of the DQC procedure caused by different data quality issues

Data quality problems	Average relative error
Missing	0.157
Sudden change	0.104
Averaging	0.074

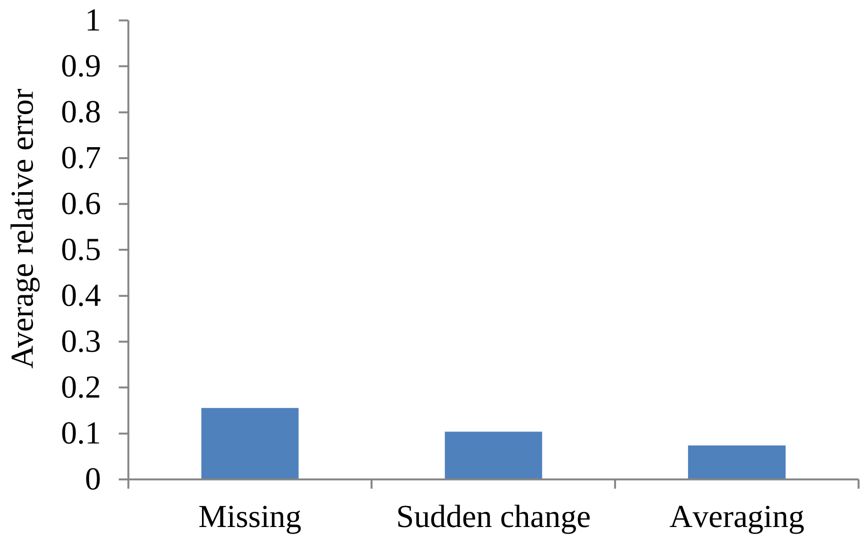


Figure 3.29: Average relative error of the DQC procedure caused by different data quality issues

From Figure 3.29 and Table 3.16, one can tell that production rates missing problem will give the worst DQC procedure result while the production rate averaging scenario has the best result in the DQC procedure.

However, all the conclusions are based on the 5x4 synthetic field where there are only 4 producers in the field. To have more assured results, a larger synthetic field with more producers is needed. In the next chapter, a 25x16 synthetic field is introduced to perform more sensitivity analysis.

Chapter 4 Data Quality Control (DQC) in 25x16 Synthetic Field

This chapter presents the validation of the DQC procedure in a large synthetic field with 25 injectors and 16 producers. This attempt has been made to produce more reliable conclusions compared to a small field with only 4 producers. Similar sensitivity analysis has been conducted to further prove the points made in Chapter 3. The interest is to discover if the DQC procedure, in a large reservoir model with different geology, behave in the same way as it is in the small synthetic field.

4.1 Introduction to the 25x16 Synthetic Field

The 25x16 synthetic reservoir is created in a commercial simulator (Eclipse E100 black oil simulator, 2005). Well spacing in the field is shown in Figure 4.1.

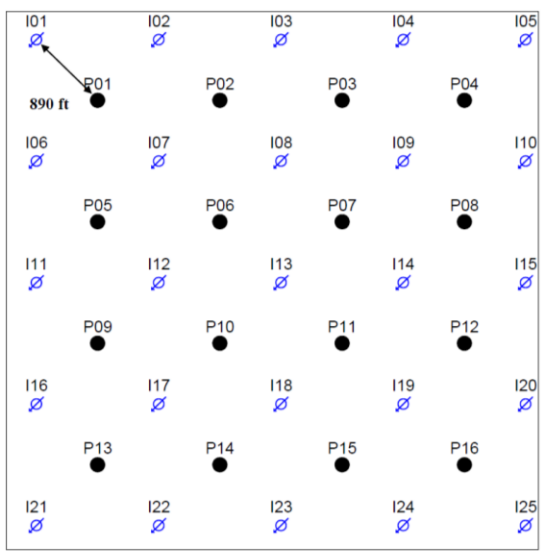


Figure 4.1: The 25x16 synthetic field well locations

The 25x16 synfield has a 5-spot injection pattern with water and undersaturated oil. General information about dimensions of the synthetic field is listed in Table 4.1.

Table 4.1: 25x16 synthetic field dimensions

Parameters	Value
Number of grids	63x63x3
Grid sizes (ft)	90x90x16
Producer-injector distance (ft)	890
Total reservoir volume (ft^3)	$1.54x10^9$

The general reservoir properties of the 25x16 synthetic field is the same as the 5x4 synthetic field (Table 3.2). In this 25x16 synfield each producer is operating under the same constant bottom-hole pressure of 250 psi. And the injection rates are fluctuating all the time. The monthly injection data (Figure 4.2) came from a real oil field and were scaled down to be proportional with the total injectivity of this synthetic case. The numerical simulation extends to 65 months, with one month for each time step.

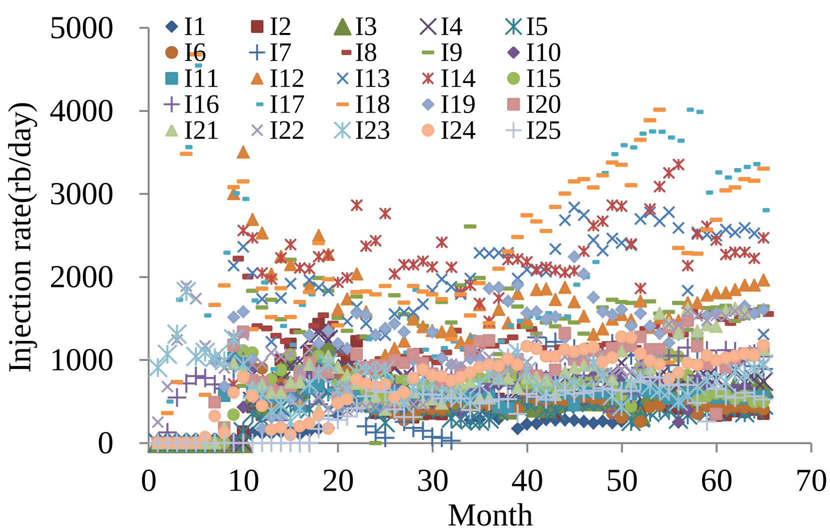


Figure 4.2 Injection rates in the 25x16 synfield

To introduce some heterogeneous characteristics to this reservoir model, we create a 1000 md high permeability square field around producer P01 and P05 (shown in Figure 4.3) while the permeability everywhere else in the field is only 40md. Further, a zero transmissibility area has been imposed in the field to separate the field into two unconnected parts as seen in Figure 4.4.

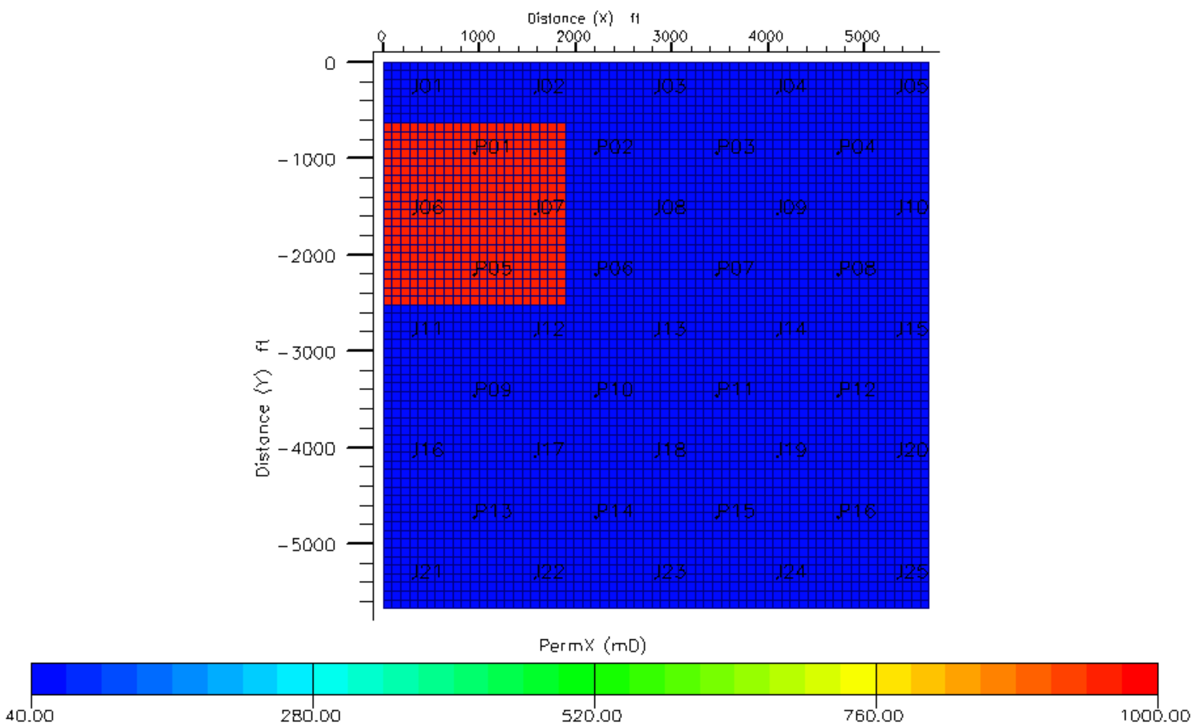


Figure 4.3: The 25x16 synfield permeability field

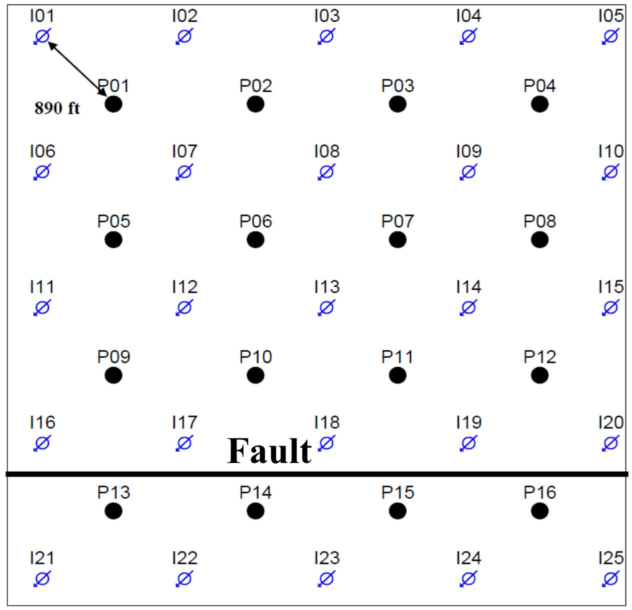


Figure 4.4: The 25x16 synfield transmissibility field

The existence of the high permeability field and the fault will add heterogeneous features to this synthetic reservoir model. After imposing injection rates, a set of simulated production rates can be obtained. These production rates are the base case of the sensitivity analysis in this chapter.

4.2 Sensitivity Analysis in the 25x16 synfield

4.2.1. Gain and Time Constant Sensitivity

In this section, again we are trying to determine if gain and time constant values will affect the DQC procedure in this large synthetic field.

The process to perform the sensitivity analysis is the same as it is in the 5x4 synthetic field. First the CRMP model will be applied to the synfield to find the gain and time constant corresponding to the uncontaminated data. Then we deliberately corrupt the production rate data in order to create different problematic data quality scenarios. After applying the DQC procedure, finally attempts are made to find out if there is a correlation between the gain (and time constant) and the quality of the DQC fits.

The time constant and connectivity after applying the CRMP model in the 25x16 synfield are listed in Figure 4.5 and Figure 4.6.

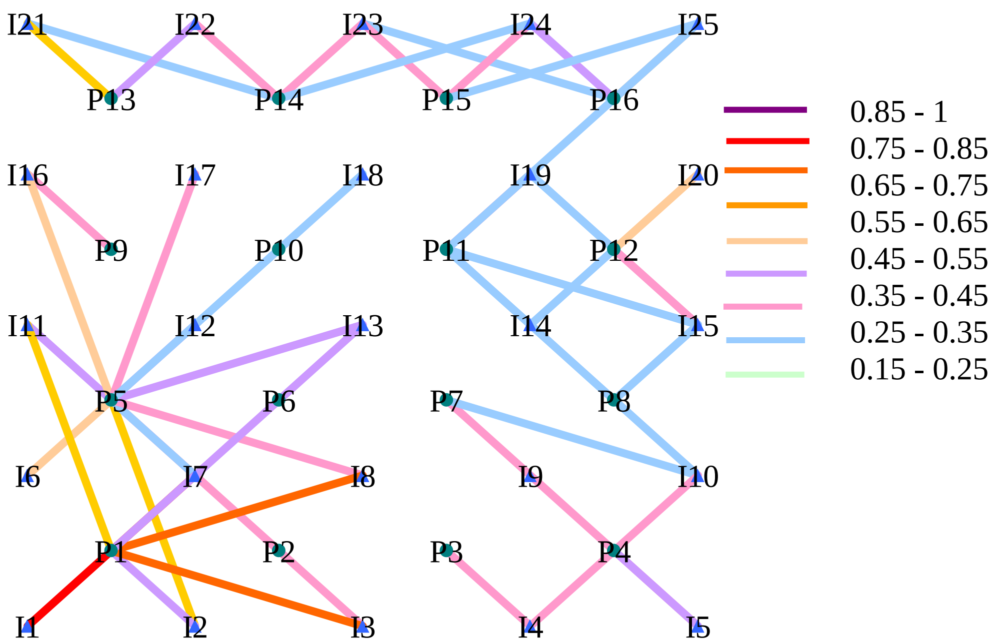


Figure 4.5: The connectivity in the 25x16 synthetic field

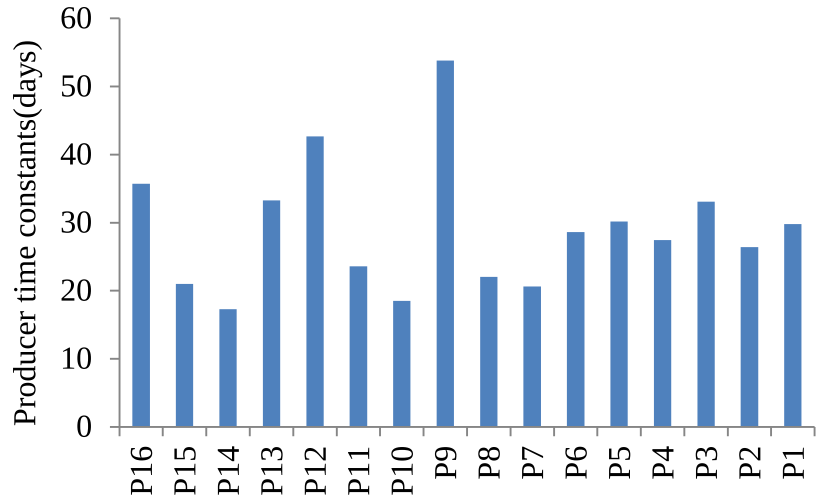


Figure 4.6: The time constants in the 25x16 synfield

However, as one can see from the connectivity map, each producer is connected to different injectors, which makes the summation of gains into a producer more meaningful than single connectivity. In other words, the sum of gains over each producer reflects the overall combined effect of the connectivity around that producer. The sum of gains values are listed in the Table 4.2 and Figure 4.7.

Figure 4.7: The sum of gains in the 25x16 synfield

Table 4.2 Time constants and sum of gains in the 25x16 synthetic field

	Producer time constant(days)	Sum of gains
P16	35.69	1.29
P15	21.00	1.09
P14	17.27	1.10
P13	33.27	1.31
P12	42.69	1.53
P11	23.57	1.60
P10	18.51	1.03
P9	53.83	0.93
P8	22.03	1.35
P7	20.63	1.36
P6	28.63	0.48
P5	30.17	4.27
P4	27.42	1.53
P3	33.08	1.04
P2	26.40	0.37
P1	29.78	4.70

From the connectivity map (Figure 4.5), one can conclude that producers P13-P16 are unconnected to the rest of the field as there is nearly no connectivity at all. However, there exists a weak connection between producer P16 and injector I19. Since it is the only connectivity across the fault, it doesn't get to change the whole picture. Besides, the CRM model is a quick reservoir simulator that, instead of requiring for geology information, gives back the characteristics of the reservoir geology. In this sense, the CRMP results shown in Figure 4.5, although not perfect, are good enough for the purpose of data quality control. From Figure 4.6, well P01 and well P05 are of high connectivity with neighbor injectors by showing high sum of gains values, which usually indicates high permeability area in the reservoir. All these results are in a good agreement with the geology of the reservoir where a fault has separated the producers P13-P16 from the rest of the field and well P01 together with well P05 are in the high permeability field of 1000md.

To test the sensitivity of the DQC process to gain and time constant, all producers have been corrupted in the same way in the same time periods to ensure the comparability. All producers have zero rates from time step 40 to 50 as seen in Figure 4.8. Applying the DQC procedure to all producers, the purpose is to compare the DQC-corrected data with the simulated data (the base case) for each producer.

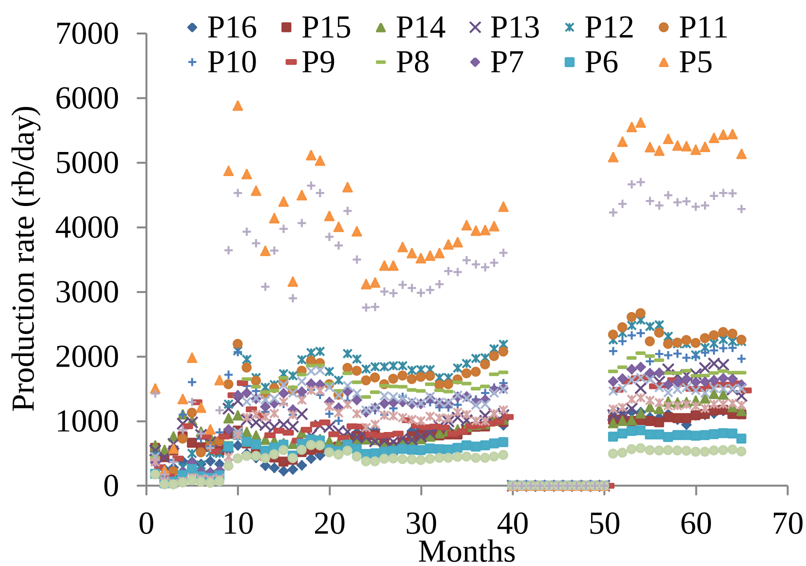
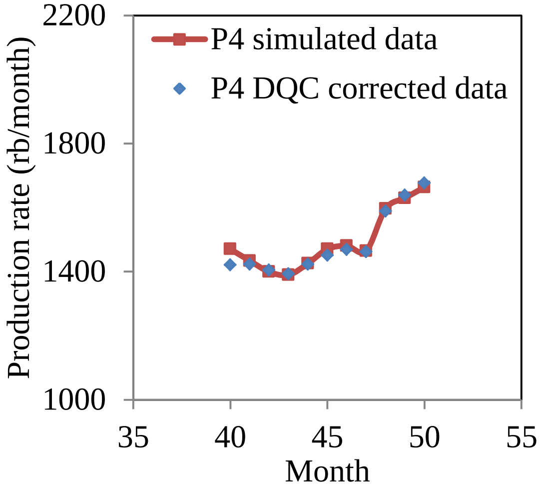
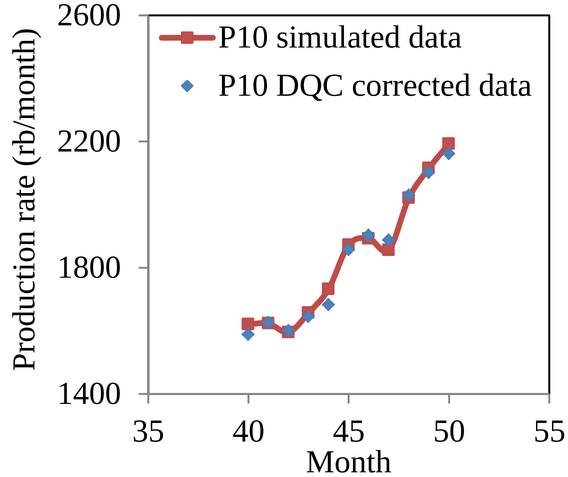


Figure 4.8: Producers with missing production rates

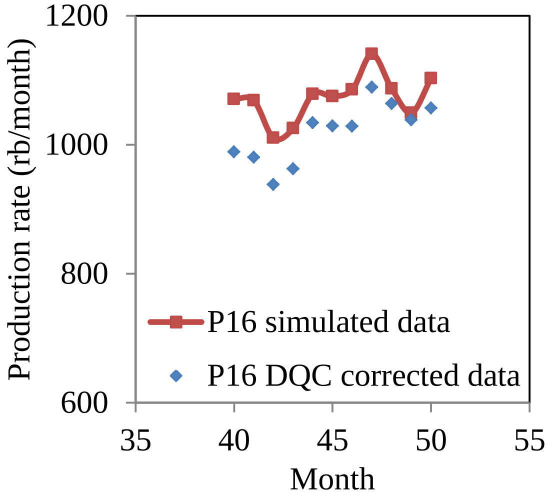
Figure 4.9 shows the matching results of simulated data vs. the DQC-corrected data for some wells.



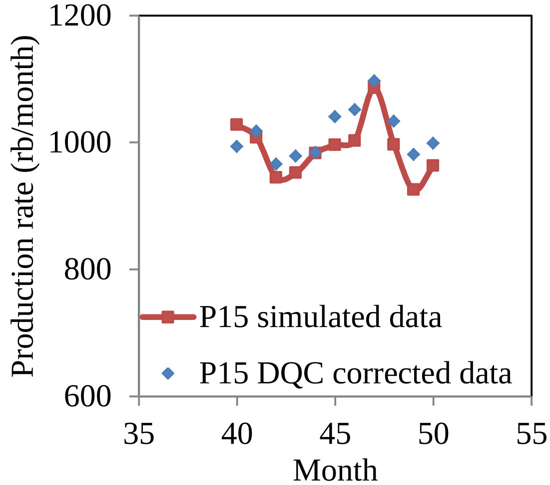
(a) Well P4



(b) Well P10



(c) Well P16



(d) Well P15

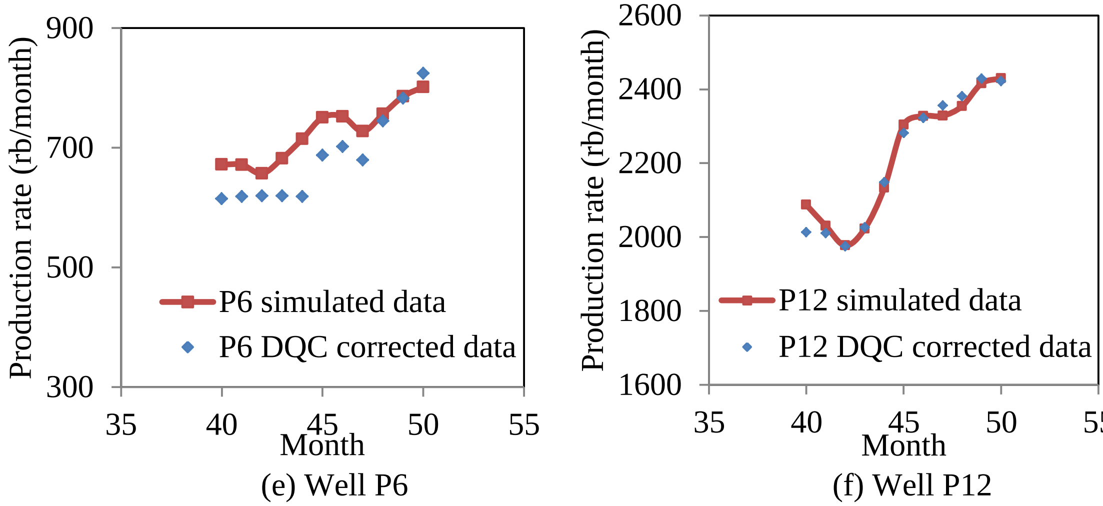


Figure 4.9: Simulated data vs. the DQC-corrected data in the gain and tau sensitivity analysis in the 25x16 synfield

Figure 4.9 reveals that some of the wells have good matching results, well P4, P10 and P12 for example. Some wells have less satisfying results, well P16, P15 and P6 for instance. The average absolute relative errors of each producer are presented in Figure 4.10.

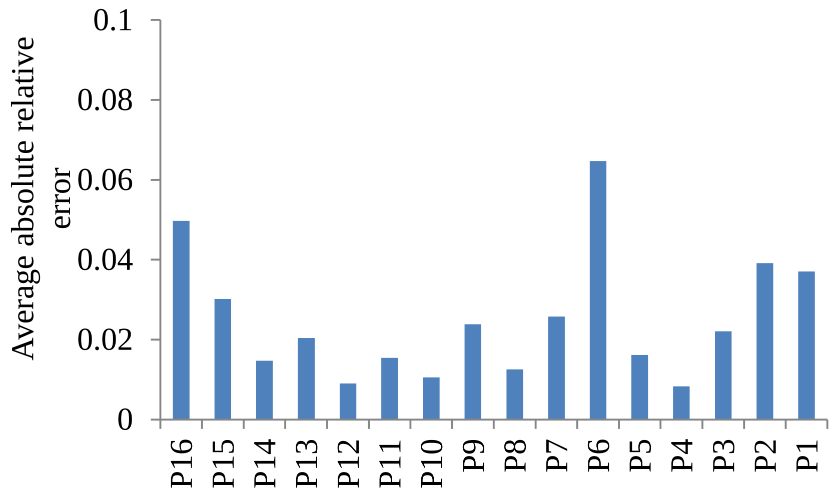


Figure 4.10: Average relative error of the DQC procedure in the gain and tau sensitivity analysis in the 25x16 synfield

In Figure 4.10, the largest average relative error is less than 0.1 which reveals that, although varied, the overall DQC fitting quality is satisfying. However, to determine the reason behind the difference in the DQC procedure fitting quality, one can tabulate the average absolute relative errors together with gain and time constant values as shown in Table 4.3:

Table 4.3: Gain and time constant with average absolute relative error in the gain and tau sensitivity analysis in the 25x16 synfield

	Average relative error	Producer time constant(days)	Sum of gain
P16	0.050	35.3	1.27
P15	0.030	26.6	1.10
P14	0.015	26.4	1.11
P13	0.020	25.9	1.28
P12	0.009	29.5	1.54
P11	0.015	22.9	1.47
P10	0.011	19.3	1.11
P9	0.024	25.6	0.94
P8	0.013	24.5	1.39
P7	0.026	22.0	1.36
P6	0.065	35.2	0.49
P5	0.016	20.5	4.61
P4	0.008	25.0	1.53
P3	0.022	26.0	1.13
P2	0.039	25.2	0.40
P1	0.037	25.8	4.18

Plots of sum of gains and time constant values versus the average absolute relative errors are shown in Figure 4.11 and 4.12.

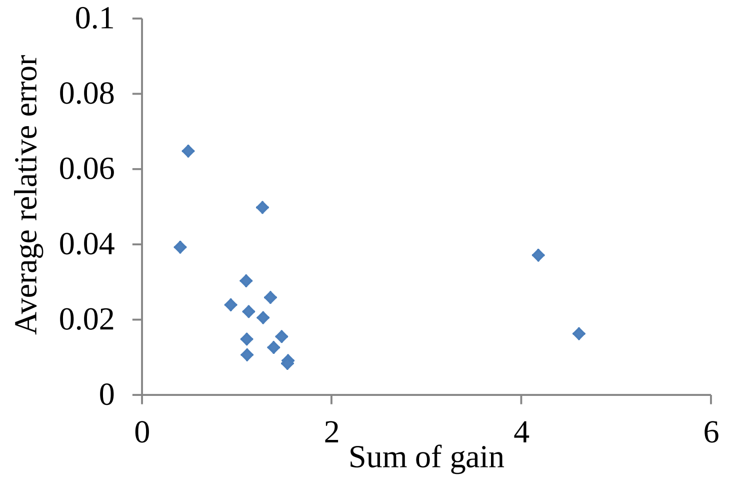


Figure 4.11: Change of relative error with sum of gain in the gain and tau sensitivity analysis in the 25x16 synfield

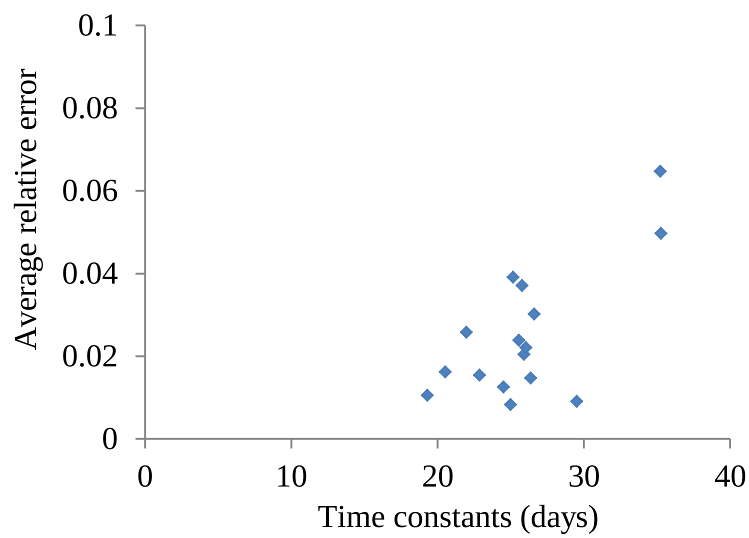


Figure 4.12: Change of relative error with time constants in the gain and tau sensitivity analysis in the 25x16 synfield

Figure 4.11 and 4.12 reveal similar information as in the smaller 5x4 synthetic field. The general trends that the higher the connectivity, the lower the average relative error and the larger the time constant, the higher the average relative error still hold in the large 25x16 synthetic reservoir. In Figure 4.11, there are two points separated far away from other points. Those two points stand for well P01 and P05, which are all in the high permeability area in the 25x16 synfield. This separation is caused by their high connectivities (sum of gains values). However, they do not have the smallest average relative errors, which conflicts with the previous conclusion we made.

This exception is caused by the CRM model itself. Although the general trend in the CRM model is that high connectivity usually comes with a small time constant while low connectivity comes with a large time constant, the CRM model doesn’t guarantee the smallest time constant for the highest connectivity. However, the time constant can affect the DQC procedure in a way that large time constant will cause an exponential-decline-like production rates so that the production response is insensitive to the injection signal. This phenomenon in return makes the DQC procedure not to work effectively. In our case, well P01 and P05, although having the largest sum of gain, don’t have the smallest time constants. This causes the DQC procedure on these two wells doesn’t give the smallest average relative errors.

Figure 4.11 and 4.12 indicate that the DQC procedure is sensitive to the value of the gain and time constant; however one can’t predict the DQC results by concluding that the highest connectivity wells will have the best DQC results. The final quality of DQC-corrected data depends on the overall combined effect of the connectivity and time constant together.

4.2.2. Number of Corrupted Data Points

The number of corrupted data points is crucial to the date quality control procedure as discussed in the previous chapter. The 80/20 rule has been introduced into the DQC procedure in Chapter 3 as a rule of thumb to provide an empirical critical percentage. Here in the 25x16 synfield, we further confirm this point.

However, unlike in the small 5x4 synfield where sensitivity analysis has only been performed on one well, all the producers in the 25x16 field will be candidates to conduct sensitivity analysis. According to Table 4.3, all producers in the 25x16 synfield have small time constants and large connectivities, which enable active responses to the injection signals.

Example 1: Fixed total number of data points, changing missing data points

In this scenario, different number of data points (0 rates) are missing while the total number of data points used remains the same (Table 4.4).

Table 4.4: Percentage of missing data points used in example 1 in number of corrupted data points sensitivity analysis in 25x16 synfield

Case	Total data points	Missing data points	Percentage of missing data points (%)
1	65	7	11
2	65	11	17
3	65	21	32
4	65	26	40
5	65	33	51
6	65	40	61

Applying the DQC procedure, the average absolute relative errors for each producer are shown in Figure 4.13. Further, the average relative errors of fitting for each situation (different percentage of missing data points) are shown in Figure 4.14.

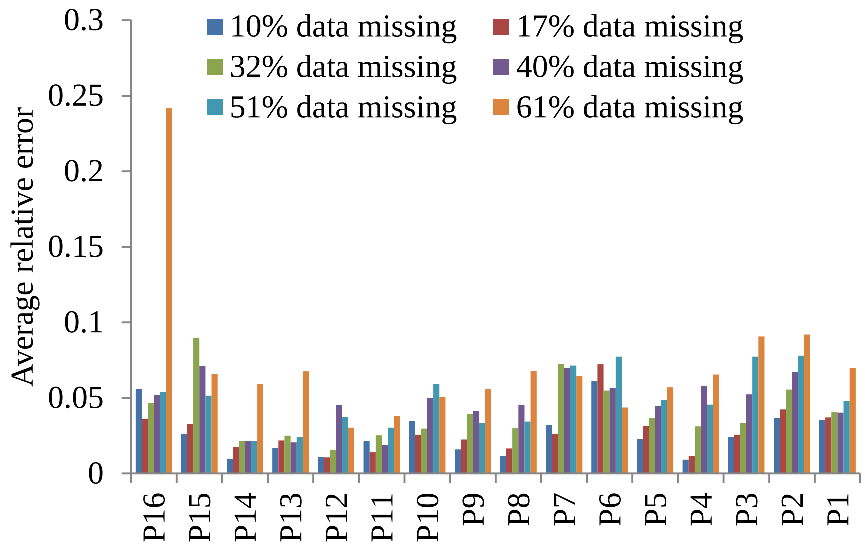


Figure 4.13: The average absolute relative error of the DQC procedure in example 1 in the 25x16 synfield

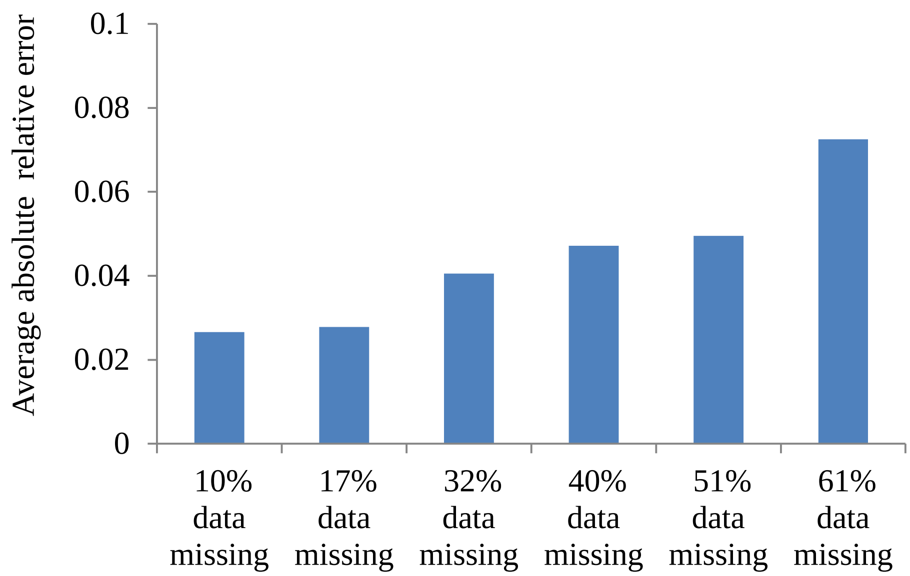
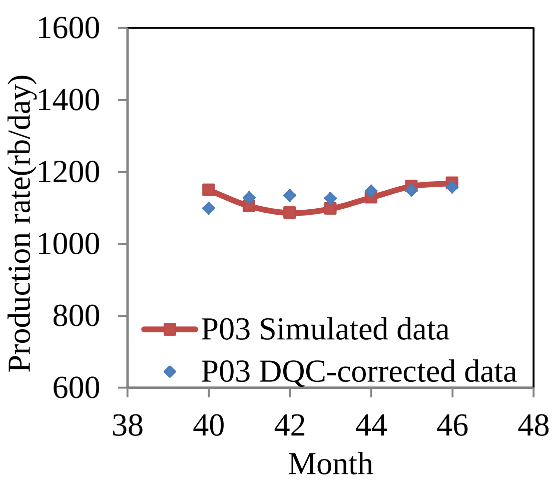


Figure 4.14 The average relative errors of the DQC procedure in example 1 in the 25x16 synfield

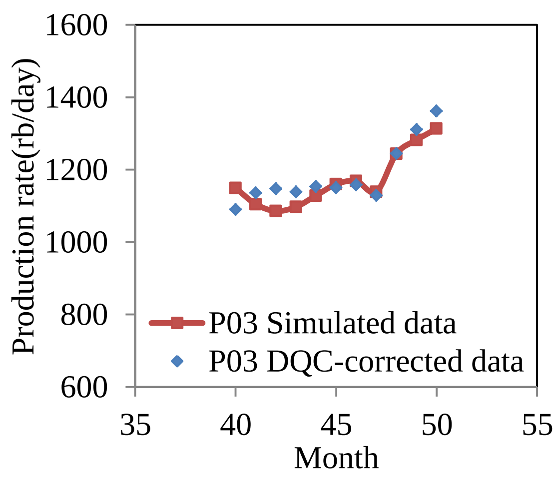
Figure 4.14 shows a clear trend that when the number of missing data points increases, the average relative errors are increased accordingly. Note that for the 7 and 11 data points missing scenarios, the average relative error is increased form 0.027 to 0.028, which is a small increase. However, the 21 data missing case has an average relative error of 0.041, which is a 45% increase compared to that of 11 months missing.

From Table 4.4, the 11 rates missing is equal to 17% of data missing, close to 20% the critical percentage mentioned in Chapter 3. The change of average relative errors in Figure 4.14 indicates that below 20% of missing data points, the average relative errors are all small as well as simliar to each other. However, above the 20% of missing data points, there is an sudden increase in the average relative errors, which futher confirms the application of the 80/20 rule as a rule of thumb in the DQC procedure.

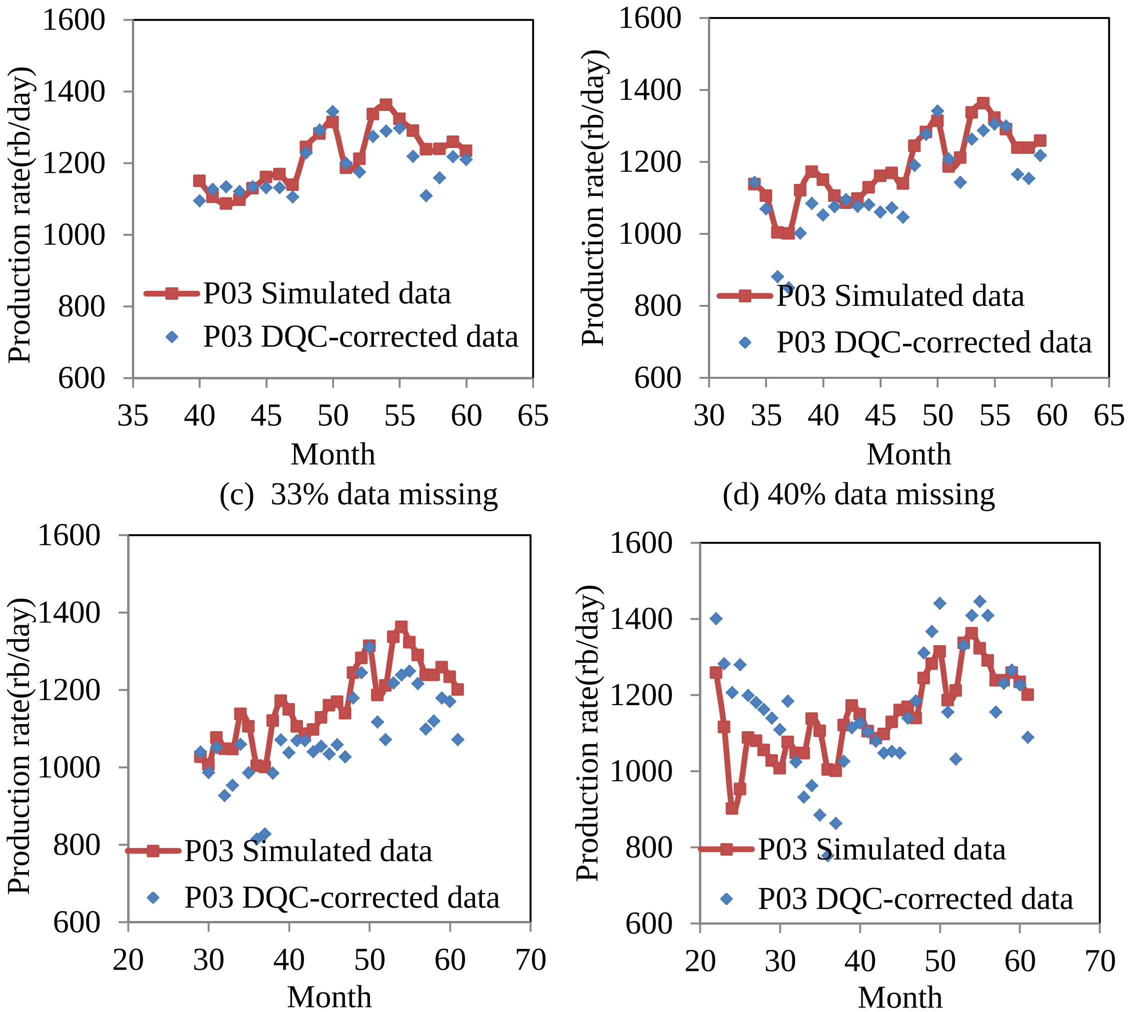
Here we show some of the matching results of some individual wells. The DQC-corrected production versus the simulated data of well P03 are showing from Figures 4.15 to 4.16.



(a) 10% data missing



(b) 17% data missing



(c) 33% data missing

(d) 40% data missing

(e) 51% data missing

(f) 61% data missing

Figure 4.15 The simulated data vs. the DQC-corrected data for well P03 in example 1 in the 25x16 synfield

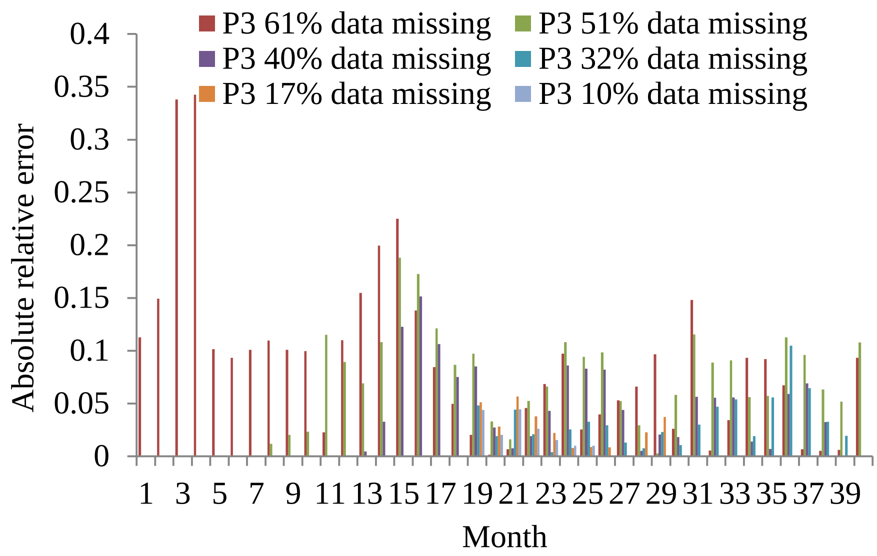


Figure 4.16 The absolute relative errors for well P03 in example 1 in the 25x16 synfield

The matching results of producer P03 under different fixed total number of data points and changing missing data points scenarios are shown in Figures 4.15 to 4.16 above. These figures show how the fitting of the DQC-corrected data to the base case get less satisfactory when we are trying to correct more data points while the total number of data point used keeps the same.

Example 2: Fixed missing data points, changing total number of data points

In this example, all producers are missing production rates (0 rates) from time step 10 to 19 which are 10 months time in total. However, the total number of data points used in this example is changing from 65 to 25 months (Table 4.6).

Table 4.5: Percentage of missing data points used in example 2

Case	Number of missing data points	Total data points	Missing data points percentage (%)
1	10	65	15
2	10	50	20
3	10	40	25
4	10	33	30
5	10	25	40

Applying the DQC procedure to each situation, the average absolute relative errors of fitting for each producer are shown in Figure 4.17. Further, the average relative errors of fitting for each situation are shown in Figure 4.18.

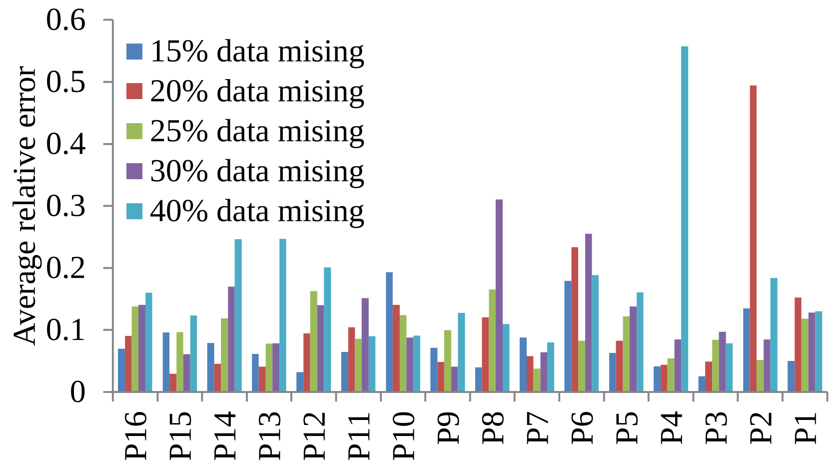


Figure 4.17: Average absolute relative errors of the DQC procedure in example 2 in 25x16 synfield

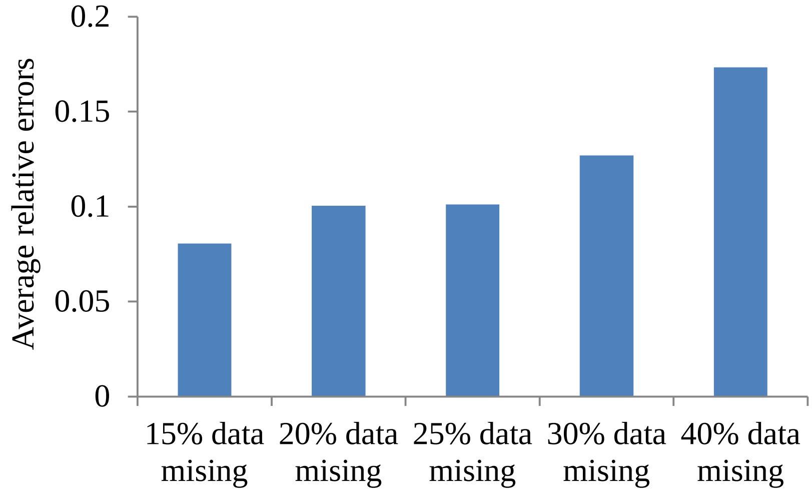
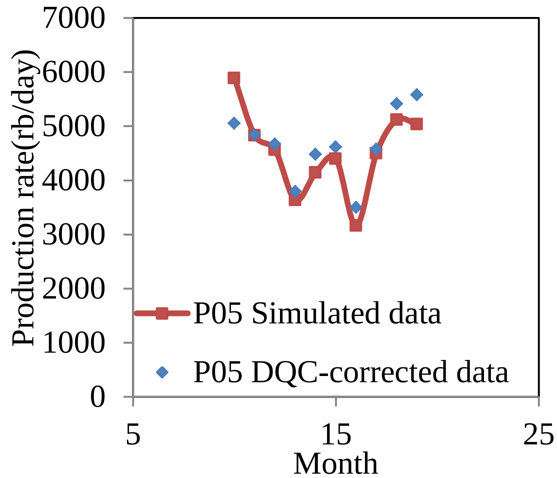
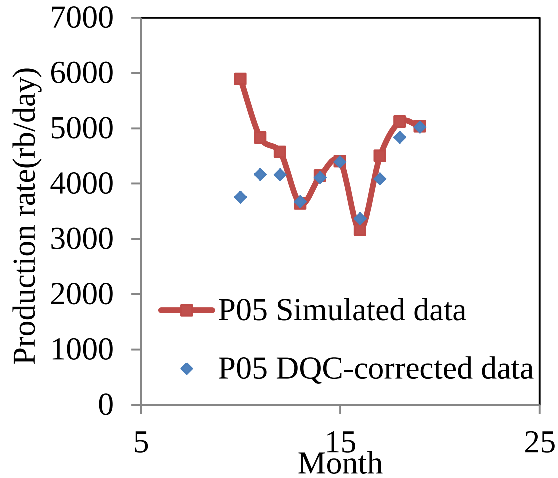


Figure 4.18: The average relative errors of the DQC procedure in example 2 in 25x16 synfield

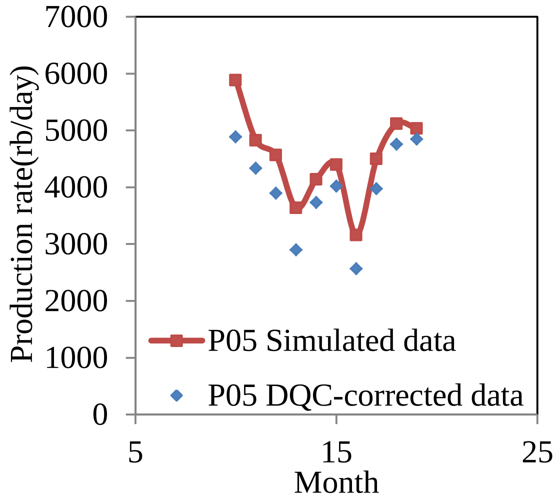
As seen from Figure 4.17, not strictly for each producer, the average absolute relative error is increased when decreasing the number of total data points used. However, Figure 4.18 tells that the overall trend is that the average relative errors of fitting is increased by increasing the percentage of corrupted data points. Here we show an example of the matching results of well P05 to give a straightforward impression of the fitting quality of the DQC-corrected data to the base case.



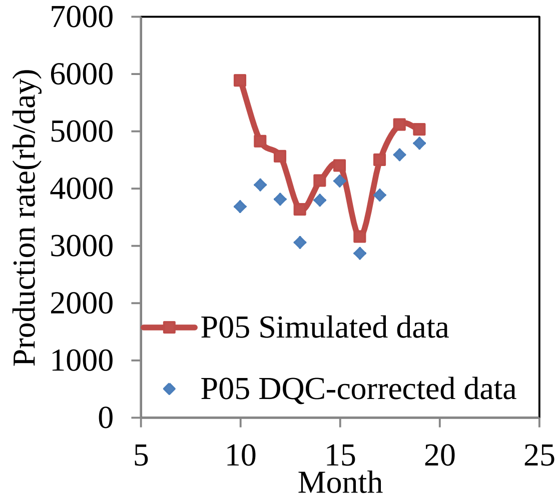
(a) 15% data missing



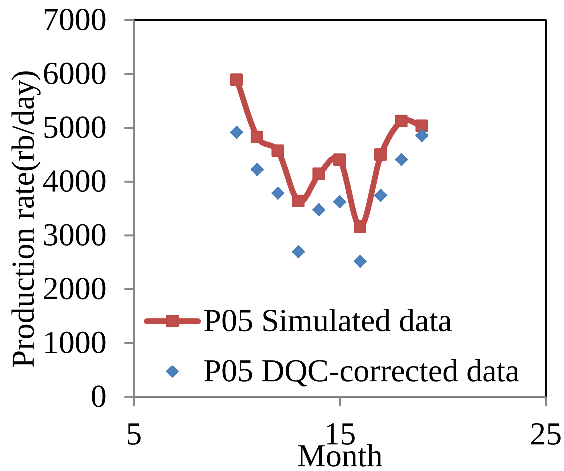
(b) 20% data missing



(c) 25% data missing



(d) 30% data missing



(e) 40% data missing

Figure 4.19 The simulated data vs. the DQC-corrected data for well P05 in example 2 in the 25x16 synfield

In example 2, 20% of missing data has an average relative error of about 10%. This number is higher than one in example 1 where 20% of missing data refers to an average relative error of about 3%. However, although not small, 10% of relative error is still acceptable. The 80/20 rule still holds. On the other hand, notice that in Figure 4.14, all scenarios have average relative errors less than 10%. When the percentage of missing data points keeps the same, it is desirable to have a case which data points are missing from a long production history rather than the case that less data points are missing from the short prodution history.

4.3.3. Different Modes of Corruption

Discussion and categorization of different types of data qulity issues have been made in the Chapter 3. In this section, sensitivity analysis is performed in the large synthetic reservoir to see if prevoius conculsions still hold. Again all the producers in the 25x16 field will be candidates to performe the sensitivity analysis.

Case A: Production rate data missing

There are 3 different scenarios of missing data:

(1) Missing data points in the front of the data set: in this scenario, all producers are missing data points (0 rates) from time step 1 to 7.

(2) Missing data points in the middle of the data set: in this case, all producers are missing data points (0 rates) from time step 40 to 46.

(3) Missing data points at the end of the data set: in this case, all producers are missing data (0 rates) from time step 54 to 60.

In other words, we deliberately create 7 data points missing in the front, middle and at the end of the data set. 7 data points missing accounts for 11% of data corruption, which based on the discussion about the critical percentage of data points missing criteria in the last section should guarantee good results. On the other hand, this means even if the DQC results are not satisfactory, it is not caused by the reason that too much data are missing.

Applying the DQC procedure, the average absolute relative errors for each producer are shown in Figure 4.20. Further, the average relative errors of fitting for each situation are shown in Figure 4.21.

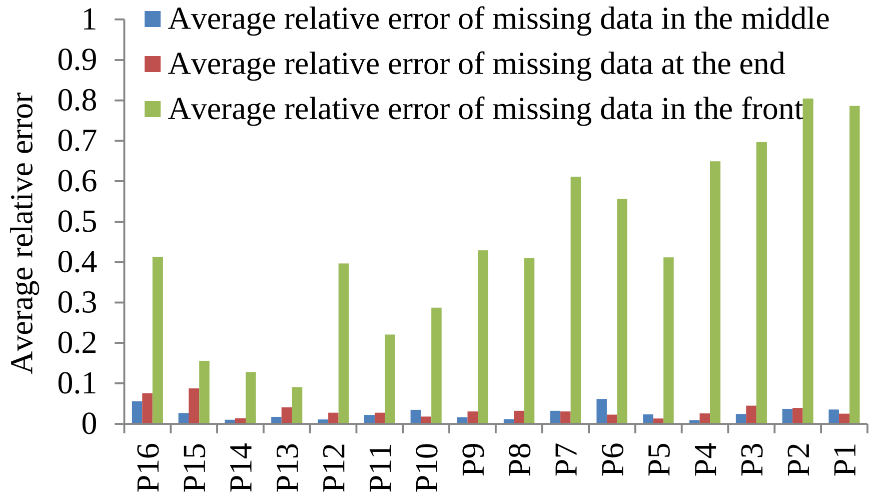


Figure 4.20: The average absolute relative error of the DQC procedure caused by data missing from different parts of the data set

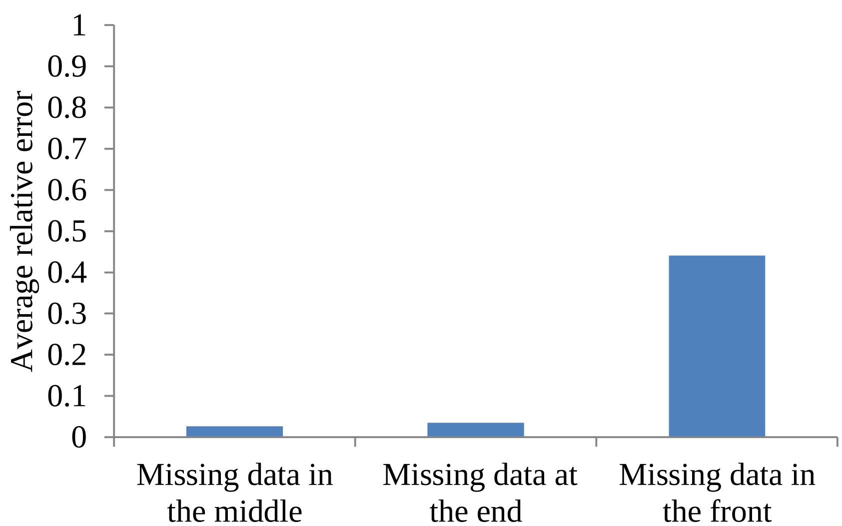


Figure 4.21: The average relative error of the DQC procedure caused by data missing from different parts of the data set

As seen from Figure 4.21, similar information has been revealed as the data missing scenario in the smaller 5x4 synthetic field. The average relative error of data missing in the front of the data set is as large as 44% while the average relative error caused by data missing in the middle and at the end of the data set are both around 3%. This indicates an overall bad DQC-correced result when data is missing from the front of the data set.

The absolute relative errors and the matching results of the DQC-corrected data vs. the base case of well P08 are shown in Figure 4.22 and 4.23 as an example to give a straightforward impression of how the fitting quality changes with position where the data missing are found.

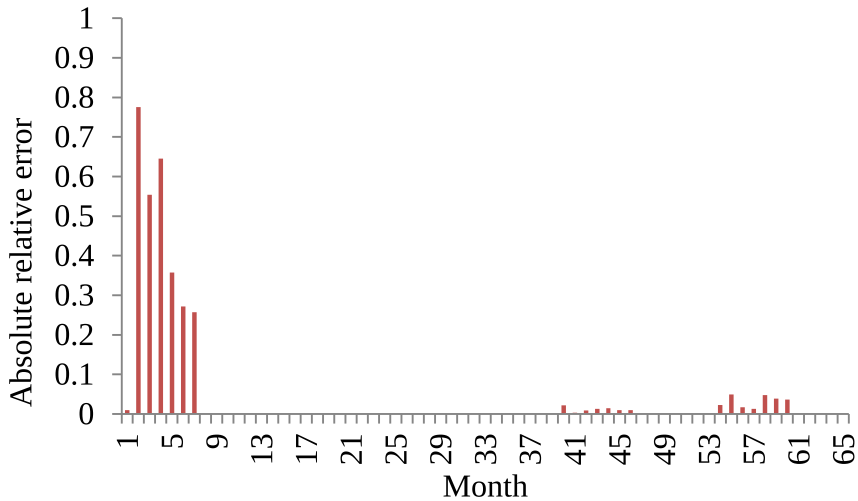
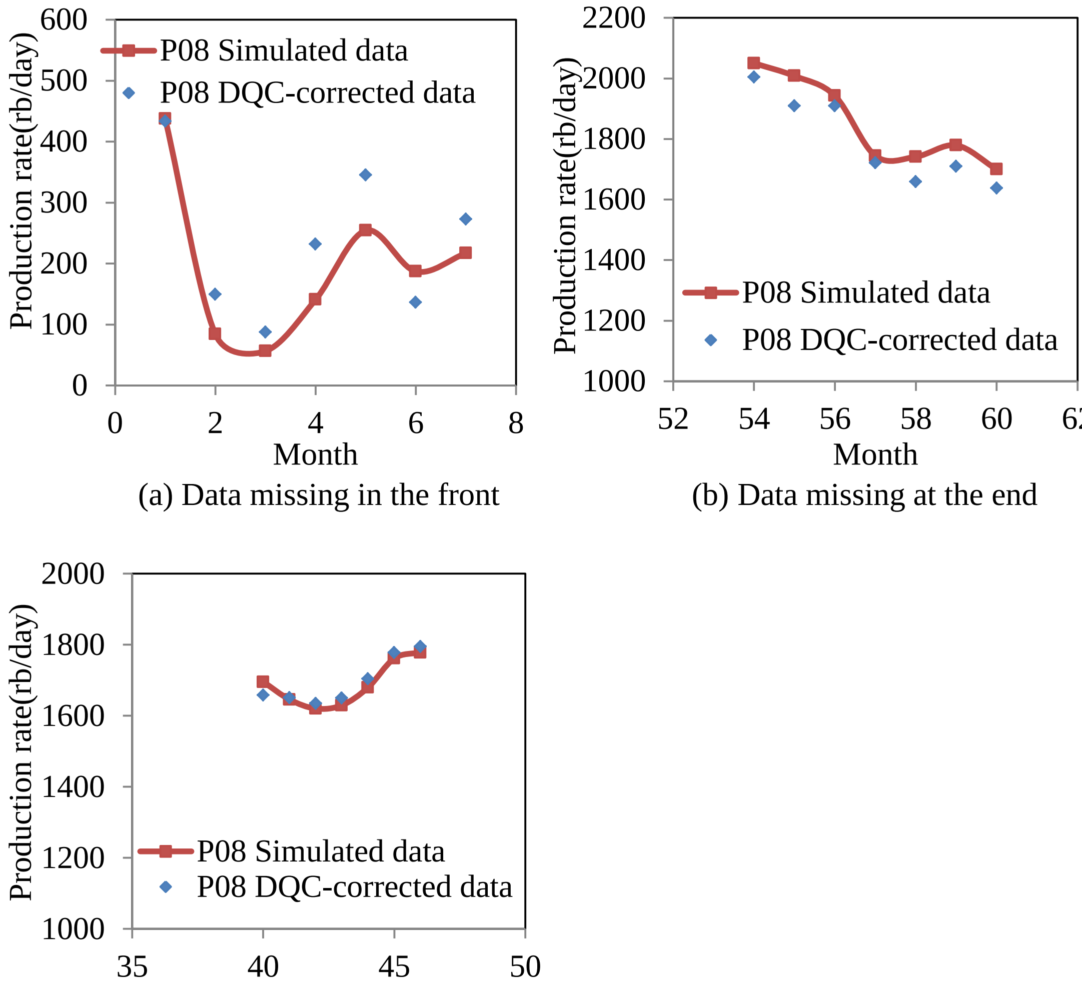


Figure 4.22: The absolute relative error of the DQC procedure caused by data missing from different parts of the data set of well P08



(a) Data missing in the front

(b) Data missing at the end

(c) Data missing in the middle

Figure 4.23: The DQC- corrected data vs. simulated data caused by data missing from different parts of the data set of well P08

Figure 4.23 shows that when data are missing from the middle and the end of the data set, the DQC-corrected results matches the original production rates impressively. However when the data is missing at the front of the data set, the DQC-corrected results are less satisfactory.

Case B: Production rate averaging

There are 3 different scenarios of data averaging:

(1) Data averaging in the front of the data set: In this case, instead of fluctuations of production rates, all producers are having a averaging rate of 548 rb/day from time step 1 to 7. The averaging rate is decided by the average production rate of all the producers in the first 7 months.

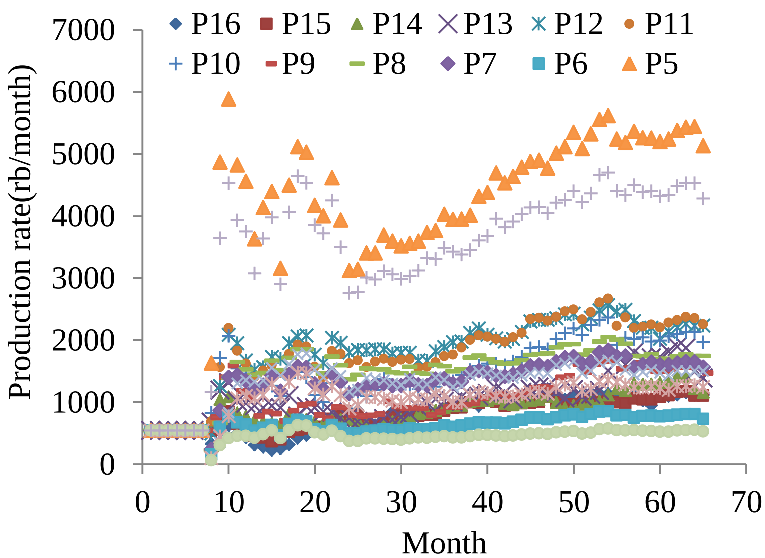


Figure 4.24: Production rate averaging at the front of the data set

(2) Date averaging in the middle of the data set: In this case, all producers are having the averaging data of 1622 rb/month time step 30 to 36. The rate here is decided by the average production rate of all the producers in the middle 7 months (time step 30 to 36).

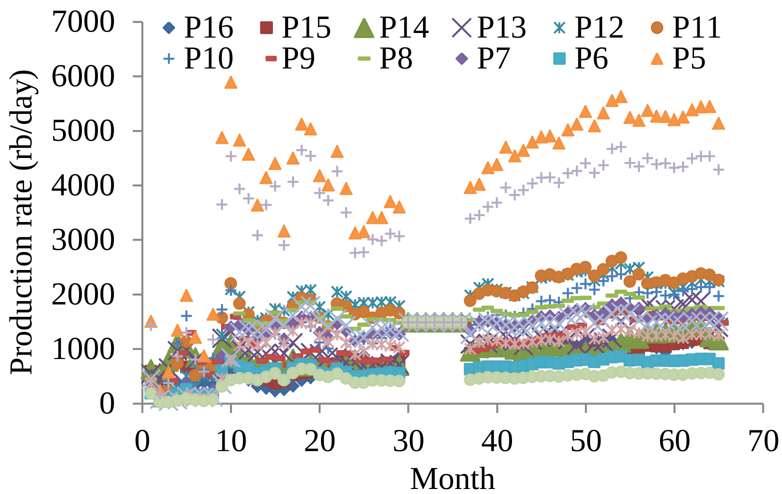


Figure 4.25: Production rate averaging in the middle of the data set

(3) Date averaging in the middle of the data set: In this case, instead of fluctuations of production rates, all producers are having the averaging data of 1989 rb/month time step 54 to 60. Again the averaging rate comes from the average production rate of all the producers in the back 7 months (time step 54 to 60).

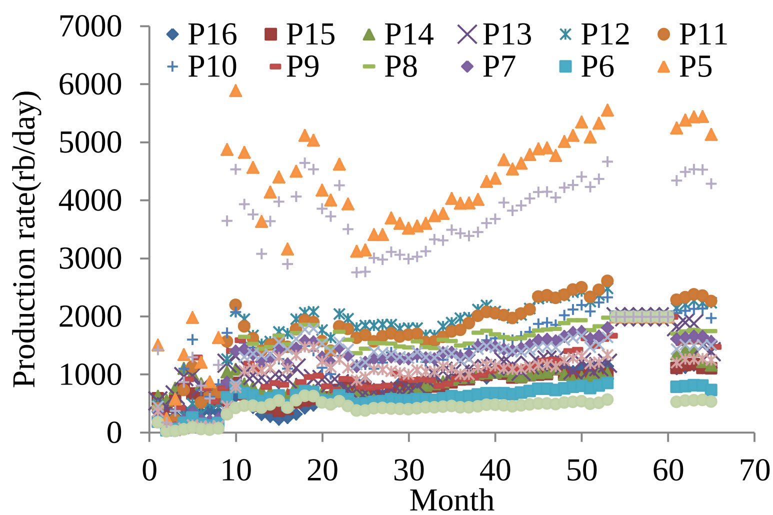


Figure 4.26: Production rate averaging at the end of the data set

The average absolute relative errors for each producer after applying the DQC procedure are shown in Figure 4.29.

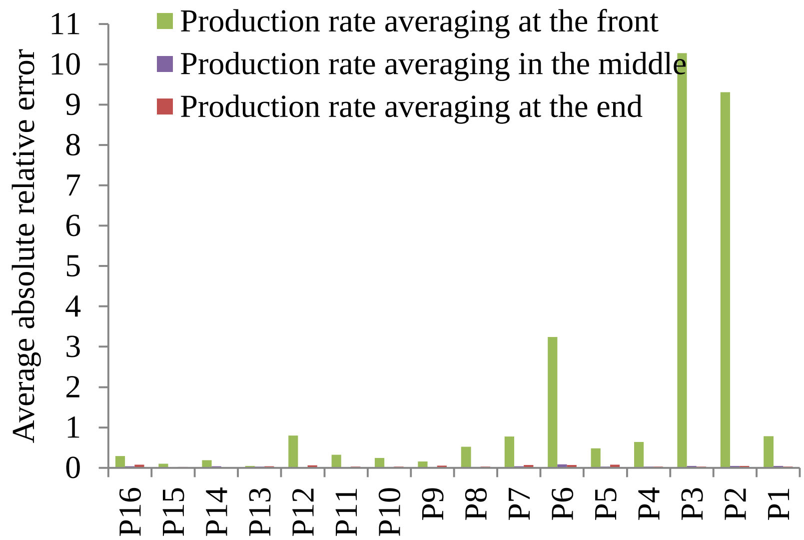


Figure 4.27: The average absolute relative error of the DQC procedure caused by the production rate averaging in different parts of the data set

In Figure 4.27, well P03 and P02 have high average absolute relative errors which appear as “spikes” in the figure above. These spikes indicate a complete failure of the DQC procedure. The average absolute relative errors of the front data averaging scenario in Figure 4.27 are way too high that the relative errors of the other two cases are barely visible. Figure 4.28 shows the average absolute relative errors of the data averaging in the middle and at the end of the data set. The largest average relative error is 0.1 in Figure 4.28.

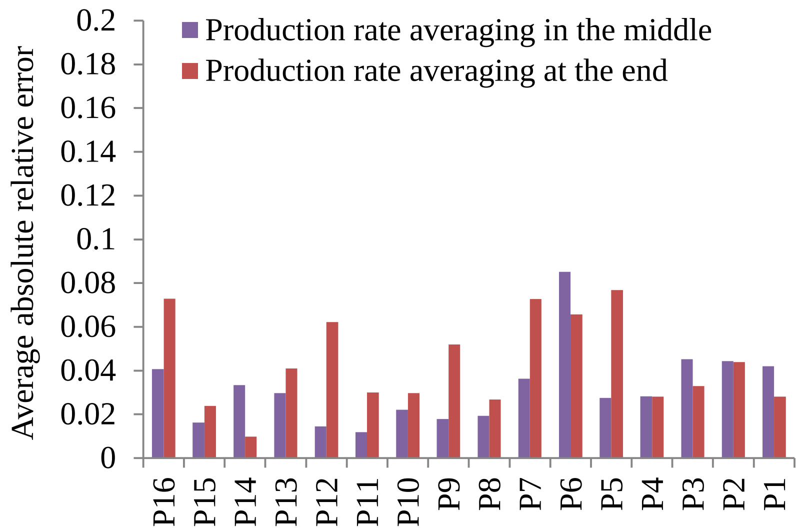


Figure 4.28: The average absolute relative error of the DQC procedure caused by production rate averaging in the middle and at the end of the data set

Further, the average relative errors of fitting for each situation are shown in Figure 4.31.

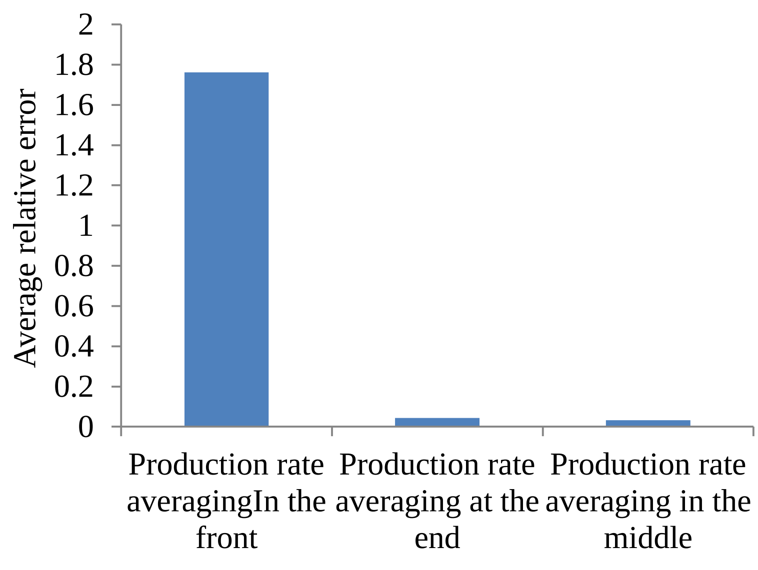


Figure 4.29: The average relative error of the DQC procedure caused by the production rate averaging in different parts of the data set

Figure 4.29 shows that, again, when data averaging happens at the front of the data set, the DQC procedure almost fails to retrieve the original data, while it will be able to give quite decent results (with average relative errors around 4%) when the averaging happens in the middle and at the end of the data set.

Here we show some matching results of individual wells. Figure 4.30 to 4.33 show the absolute relative errors for each month and the matching results for well P07 and P06. We put the matching results of the simulated rates vs. the DQC corrected rate under different production data averaging scenarios in the same plot for well P06 and P07 respectively to show , straightforwardly, the difference in fit quality.

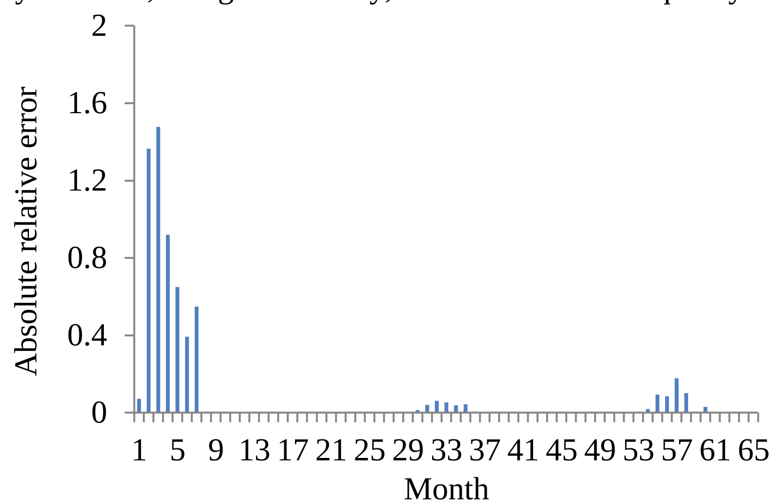


Figure 4.30: The absolute relative error of the DQC procedure caused by production rate averaging in different parts of the data set of well P07

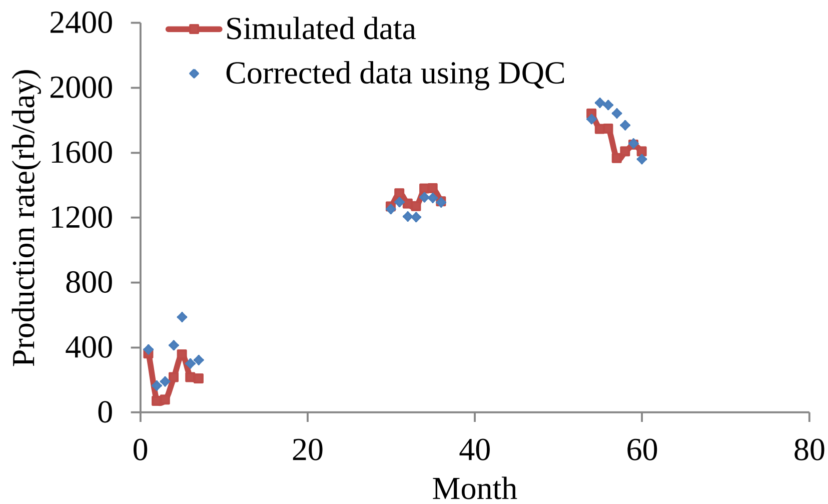


Figure 4.31: The fitting of corrected data vs. simulated data of the DQC procedure caused by production rate averaging in different parts of the data set of well P07

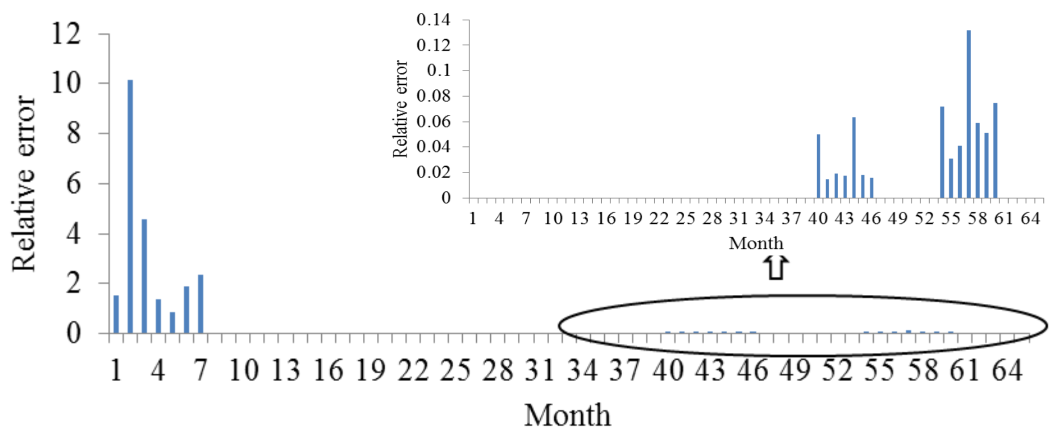


Figure 4.32: The absolute relative error of the DQC procedure caused by production rate averaging in different parts of the data set of well P06

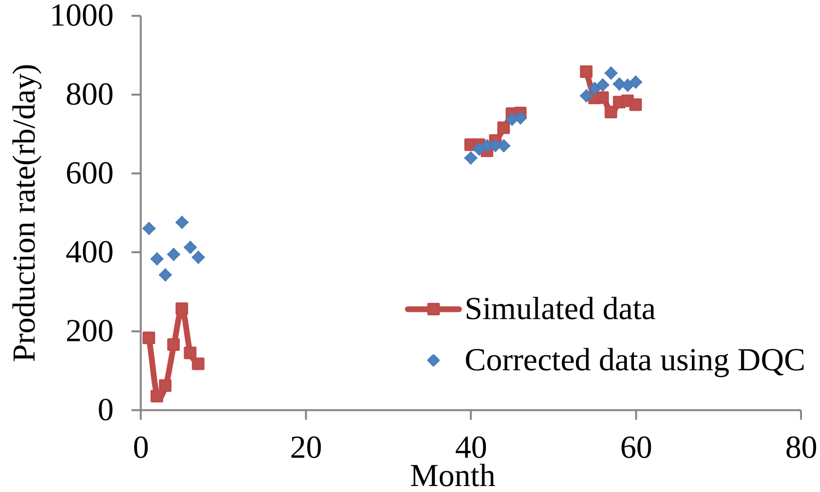


Figure 4.33: The fitting of corrected data vs. simulated data of the DQC procedure caused by production rate averaging in different parts of the data set of well P06

These figures show directly how the quality of the DQC-corrected data changes when date averaging scenario happens in different parts of the data set. We also have conducted a sensitivity analysis to decide if the averaging value itself would affect the DQC procedure. The conclusion reached is the same as that in the small 5x4 synfield: the DQC procedure is sensitive to where the data averaging happens rather than the averaging value itself.

Case C: Production rate sudden change

There are 3 different scenarios:

(1) Sudden rate change in the front of the data set: In this case, instead of normal fluctuations of production rates, all producers have a jump in the data, either an increase or a decrease.

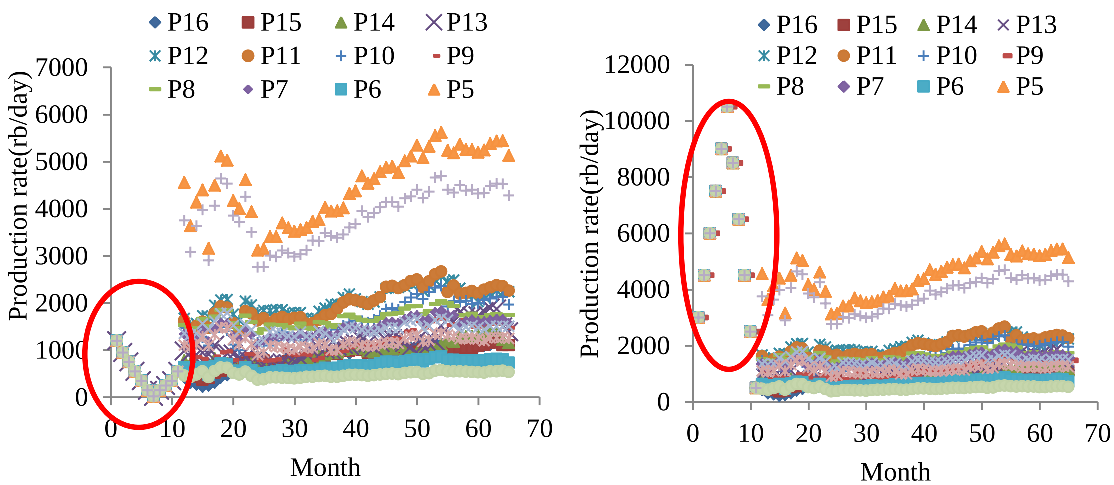


Figure 4.34: Production rate sudden change in the front of the data set

(2) Sudden rate change in the middle of the data set: all producers have a jump in the data, either an increase or a decrease, in the middle of the data set.

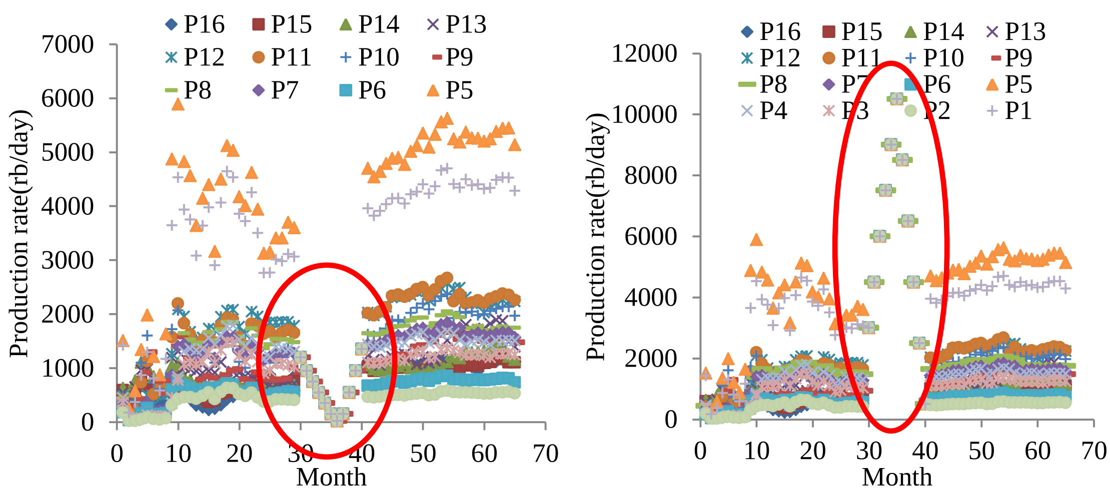


Figure 4.35: Production rate sudden change in the middle of the data set

(3) Sudden rate change at the end of the data set: all producers have a jump in the data, either an increase or a decrease, in the back of the data set.

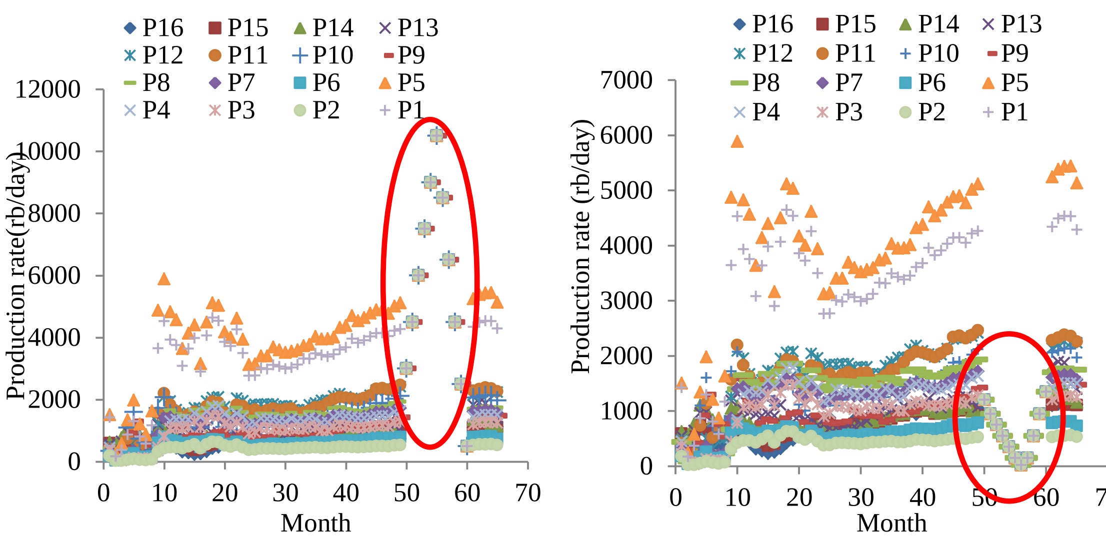


Figure 4.36: Production rate sudden change at the end of the data set

The average relative error for each production rate sudden change situation after applying the DQC procedure are shown in Figure 4.39.

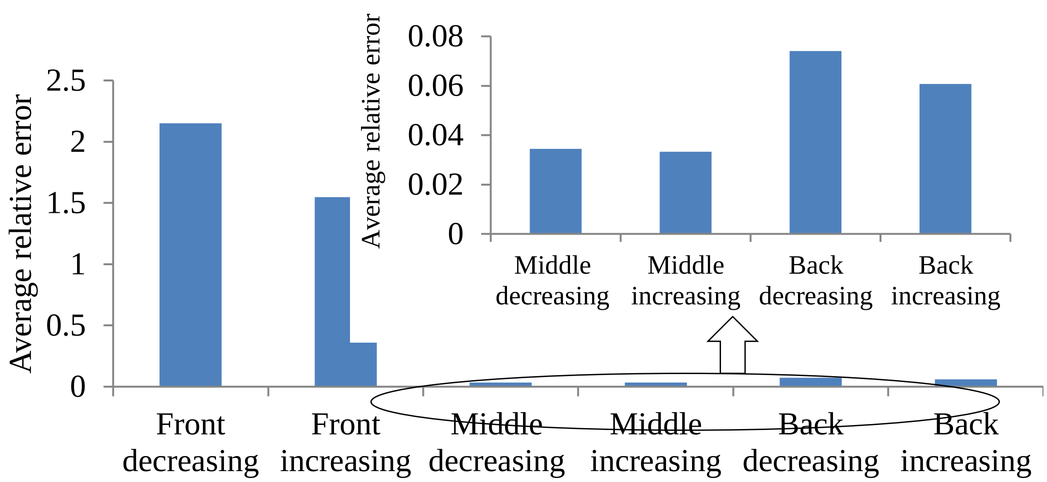
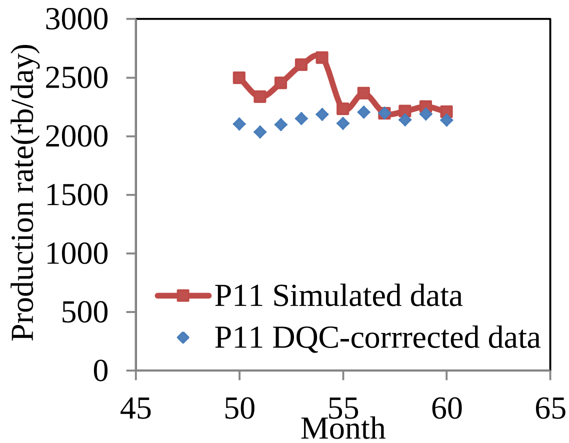
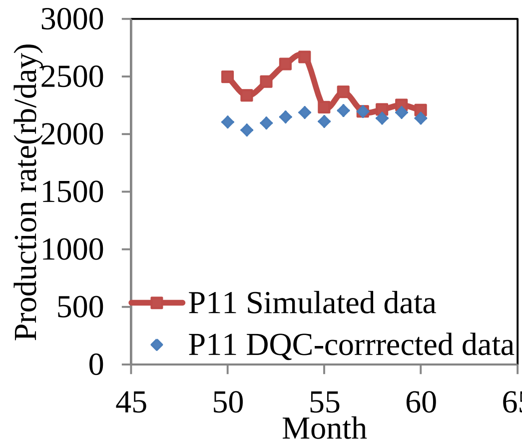


Figure 4.37: Average relative error of the DQC procedure caused by different types of production rates sudden change

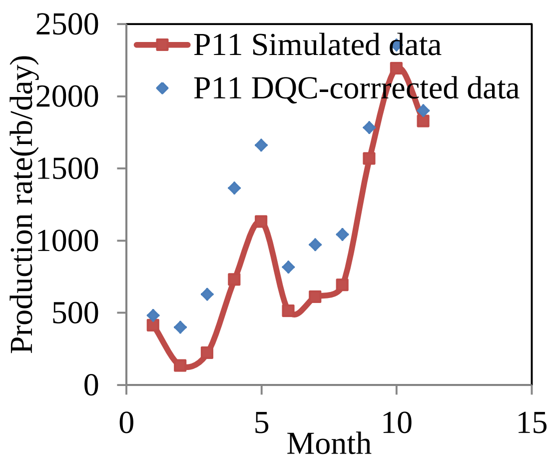
Figure 4.37 shows that whether sudden rates decrese or increase do affect the results of the DQC procedure. However, when sudden rates change is in the middle or at the end of the data set, the average relative errors coresponding to increase and decease, although different, are all quite small. Figure 4.38 shows matching results for the production rate sudden change scenarios for well P07. Comparing the right and left side in Figure 4.38, one can hardly tell the difference of fits when it comes to sudden rates change in the middle or at the end of the data set.



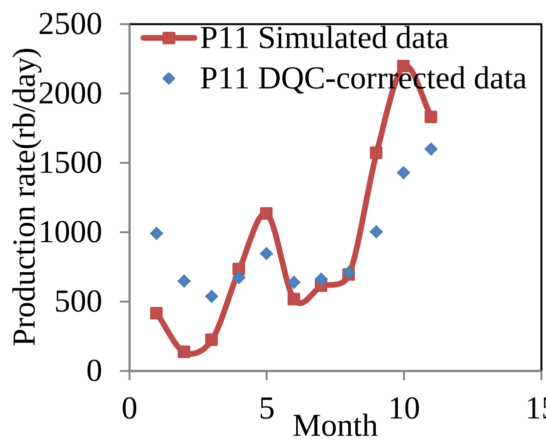
(a) Back increase



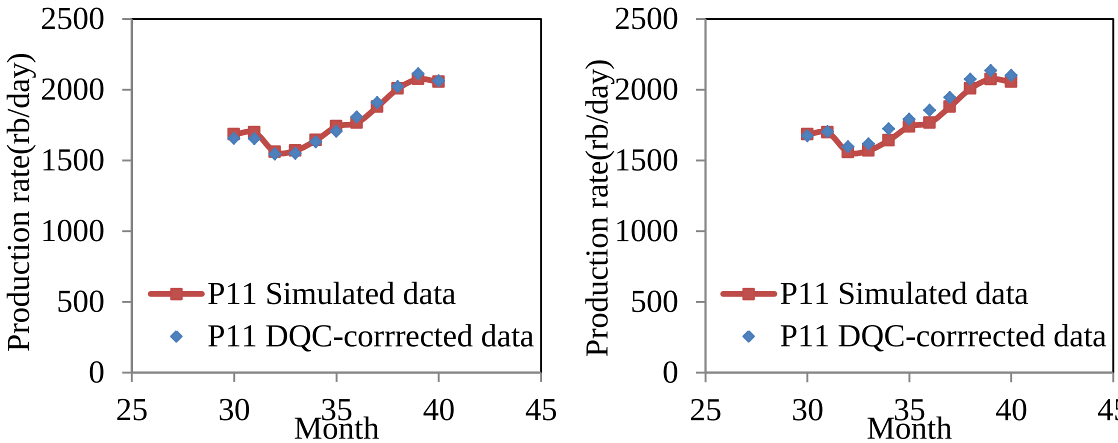
(b) Back decease



(c) Front increase



(d) Front decease



(e) Middle increase

(f) Middle decease

Figure 4.38: The fitting of corrected data vs. simulated data of production rate averaging of well P11

4.3 Summary and Conclusion

In this chapter, we discussed the sensitity analysis in the 25x16 sythetic field. we attempt to prove the conclusions about this new method made in Chapter 3. The interest is to discover if the DQC procedure, in a large reservoir model with different geology, behave in the same way as it is in the small synthetic field.

1. Gain and time constants

In general, the conclusion made in Chapter 3 holds in the large 25x16 synthetic reservoir model. High connectivity still comes with a better DQC-corrected result while large time constant often gives less satisfying DQC-retrieved result. However the CRM model doesn't guarantee that a well with the highest connectivity will necessarily have the smallest time constants. Large time constant will cause an exponential-decline-like production rate making the production response not sensitive to the injection signal, which will in return make the DQC procedure not to work effectively. Thus one can't conclude that that high connectivity wells would have the best DQC-corrected results. The final quality of DQC-retrieved data depends on the overall combined effect of connectivity and time constant together.

2. Number of corruption periods

We test the 80/20 rule in the 25x16 large synfield. The results show that below 20% of data corruption, the average relative errors are all small and similar, while if higher than the 20% critical percentage, the average relative errors can be large and can increase rapidly.

3. Different modes of corruption

As we discussed in Chapter 3, the main goal to conduct sensivity analysis to test different modes of corruption is to decide where to put the problemetic data to get the best DQC-corrected results and under which mode of corruption that the DQC procedure works the best.

(1) Summarize by the position where the problematic data appears in the data set: Table 4.6, Table 4.7 and Table 4.8 summarize the average relative errors of each mode of data corruption in the front, middle and at the end of the data sets.

Table 4.6 Average relative error of different data quality issues in the front of the data set

Data quality problems in the front	Average relative error
Missing	0.44
Averaging	1.76
Sudden decrease	2.15
Sudden increase	1.55

Table 4.7 Average relative error of different data quality issues in the middle of the data set

Data quality problems in the middle	Average relative error
Missing	0.027
Averaging	0.032
Sudden decrease	0.034
Sudden increase	0.033

Table 4.8 Average relative error of different data quality issues at the end of the data set

Data quality problems at the end	Average relative error
Missing	0.034
Averaging	0.044
Sudden decrease	0.074
Sudden increase	0.061

Table 4.9 Average relative error at different positions in the data set

Position in the data set	Average relative error
Front	1.47
Middle	0.03
End	0.05

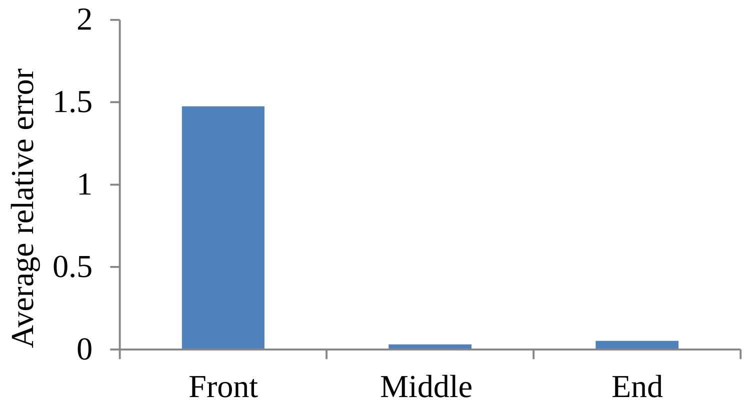


Figure 4.39 Average relative error of each part in the data set

Table 4.9 and Figure 4.39 reveal that, no matter which data qulity problem it is, the middle part of the data set is the position where always has good quality of DQC-corrected results (the average raltive error equals to 0.03). Compare this result with the one in the 5x4 synfield (Table 3.14), the average relative error of all data quality problems in the middle is also 0.03. Compared to the smaller 5x4 synthetic field, the average relative errors of all data quality problems in the front of the data set increased a much more and that in the back of the data set has decreased by 50%.

(2) Summarise by the different modes of corruption : summrize the average relative errors with respect to the fashion of corruption in Figure 4.40 and Table 4.10 below:

Table 4.10 Average relative error of different data quality issues

Data quality problems	Average relative error
Missing	0.167
Averaging	0.032
Sudden change	0.061

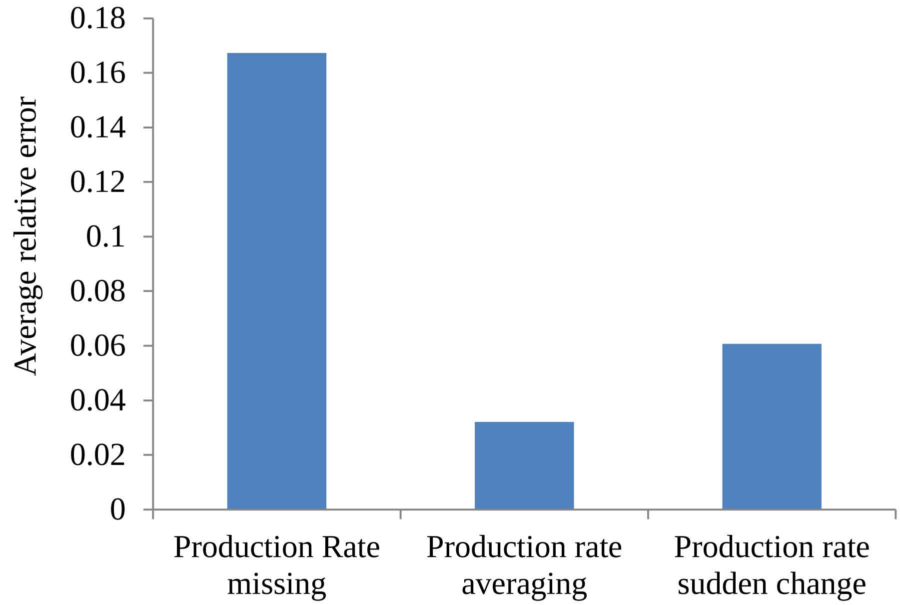


Figure 4.40: Average relative error of different data quality problems

Once again the data missing issure gives the worst DQC-correced results as it does in the smaller 5x4 synthetic field. And data averaging mode, again, will give the best DQC-retrieved data just like the 5x4 synthetic field. Comparing the resluts shown in Table 4.10 with Table 3.16, the average relative errors of data missing and sudden rates change scenarios are almost the same. However the the average relative error of data averaging is even lower in the large synthetic field.

Chapter 5: Summary, Conclusion and Future Recommendation

The purpose of this thesis is to show that the data quality control procedure in the CRM model can be used to retrieve production data as it was before any mistake or corruption. To this end, methods of the DQC procedure were developed and used to clean reservoir production data.

This chapter presents a summary of the technical contributions of this work, the conclusions reached, and recommendations for future research in this area.

5.1 Summary and Conclusion

Production data are the most abundant data in the field. However, they can often be of poor quality because of undocumented operational problems, or changes in operating conditions, or even recording mistakes. To restore issue-free production data, this thesis introduced a new method to control data quality. We use the CRM model itself as a data quality control tool. CRM is a simple reservoir simulation model that characterizes the connectivity between injector and producers using only production and injection data. Because the CRM model is based on the continuity equation, it can be used to analyze the production corresponding to the injection signal in the reservoir. The problematic production data are then put into the CRM model directly and the resulting CRM output parameters are used to evaluate what the correct production response would be under current injection scheme.

Sensitivity analysis has been conducted in different sizes of synthetic fields, which are heterogeneous ideal reservoirs with imposed geology and well features in

Eclipse. The aim is to show how bad data could be misleading and the best way to restore the production data.

Based on the sensitity analysis in sythetic fields, we attempt to reach some conclusions about this new theory.

1. Gain and time constants

In general, the conclusions made in both Chapter 3 and Chapter 4 are consistent. High connectivity still comes with a better DQC-corrected result while large time constant often gives less satisfying DQC-retrieved results. However the CRM model doesn't guarantee that a well with the highest connectivity will necessarily have the smallest time constants. Large time constant will cause an exponential-decline-like production rate making the production response insensitive to the injection signal, which will in turn make the DQC procedure work poorly. Thus one can't conclude that that high connectivity wells would have the best DQC-corrected results. The final quality of DQC-retrieved data depends on the overall combined effect of connectivity and time constant.

2. Number of corruption periods

For the purpose of data quality control, the 80/20 rule is used as a rule of thumb. It means that 20% is a critical percentage of corrupted data. In other words, 20% or less bad effect is acceptable in the DQC procedure. The other 80% must be good data for the method to work. By testing the 80/20 rule in two different synthetic reservoirs, we are confident that when less than 20% of data is corrupted, the DQC procedure is able to give back good quality of corrected data.

3. Different modes of corruption

The main goal to conduct sensivity analysis to different modes of corruption is to decide where to put the problemetic data to get the best DQC results and under which mode of corruption the DQC procedure works the best. The sensivity analysis results in 5x4 synthetic field and 25x16 synthetic field are in a good agreement. The middle of a data set is the position where always shows good quality of DQC-corrected results. Thus it is desireable for the problematic data to be in the middle of the data set. However, if we can't choose where and how the data is corrupted, and the sensivity analysis would give a overall confidence limit to expect when it comes to different corruption modes.

5.2 Future Work

The following future work is recommended in regards to the data quality control procedure:

1. Applying the DQC to synthetic fields with more complex geology to further verify that the DQC can be used effectively regardless of the geological complexity.

2. The cases studied in this thesis only account for obviously problematic production data; however injection data can also be problematic. Further, extending the DQC procedure to injection data could be a meaningful future work.

3. This study was restricted to water floods, but there is no reason that the CRM could not be extended to other types of enhanced oil recovery. Extending the CRM model, as well as the DQC procedure in gas, chemical and thermal floods, etc. would be recommended.

4. Since the production data quality problem can be caused by many reasons, communication with production or operation engineers is very important. The

involvement of field engineers in the development the data quality control task is highly recommended.

References

Albertoni, A. 2002. Inferring Interwell Connectivity Only From Well-Rate Fluctuations in Waterfloods. M.S. Thesis, The University of Texas at Austin, Austin, Texas.

Albertoni, A. and Lake, L. W. 2003. Inferring Connectivity Only From Well-Rate Fluctuations in Waterfloods. SPE Reservoir Evaluation and Engineering Journal, 6 (1): 6-16.

Agarwal, R.G., Gardner, D.C., Kleinsteiber, S.W. et al. 1999. Analyzing Well Production Data Using Combined-Type-Curve and Decline-Curve Analysis Concepts. SPE Reservoir Evaluation & Engineering 2 (5): 478-486. DOI: 10.2118/57916-pa

Al-Habsi, M. and Stoffels, P. 2005. Intensive Data Gathering through a Waterflood Pilot for Redevelopment of a Giant Fractured Carbonate Field, Oman. Paper presented at the International Petroleum Technology Conference, Doha, Qatar. International Petroleum Technology Conference 10794-MS.

Bell, M.C. and Fisk, J.H. 2008. Data Access and Integration for Major Capital Projects: A Case Study. Paper presented at the Intelligent Energy Conference and Exhibition, Amsterdam, The Netherlands. Society of Petroleum Engineers 112203-MS.

Ershaghi, I., Ortega, A., Lee, K.-H. et al. 2008. A Method for Characterization of Flow Units between Injection-Production Wells Using Performance Data. Paper presented at the SPE Western Regional and Pacific Section AAPG Joint Meeting, Bakersfield, California, USA. Society of Petroleum Engineers 114222-MS.

Faskhoodi, M.M. and Kelkar, M.G. 2007. Using the Low Frequency Observation Data in

Automatic History Matching. Paper presented at the SPE Middle East Oil and Gas Show and Conference, Kingdom of Bahrain. Society of Petroleum Engineers 105252-MS.

Gentil, P. H. 2005. The Use of Multilinear Regression Models in Patterned Waterfloods: Physical Meaning of the Regression Coefficients. M.S. Thesis, The University of Texas at Austin, Austin, Texas.

Kabir, C.S. and Young, N.J. 2001. Handling Production-Data Uncertainty in History Matching: The Meren Reservoir Case Study. Paper presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana. Copyright 2001, Society of Petroleum Engineers Inc. 71621-MS.

Kaviani, D., Jensen, J.L., Lake, L.W. et al. 2008. Estimation of Interwell Connectivity in the Case of Fluctuating Bottomhole Pressures. Paper presented at the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE. Society of Petroleum Engineers 117856-MS.

Khasanov, M., Krasnov, V., and Guk, V. 2008. Reservoir Parameters Evaluation Based on Production Data Analysis (Russian). Paper presented at the SPE Russian Oil and Gas Technical Conference and Exhibition, Moscow, Russia. Society of Petroleum Engineers 117406-RU. DOI: 10.2118/117406-ru.

Lake, L.W. 1989. Enhanced Oil Recovery. Prentice Hall, Englewood Cliffs, New Jersey

Lake, L.W., Liang, X., Edgar, T.F. et al. 2007. Optimization of Oil Production Based on a Capacitance Model of Production and Injection Rates. Paper presented at the Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas, U.S.A. Society of Petroleum Engineers 107713-MS.

Liang, X., Weber, D.B., Edgar, T.F., Lake, L.W., Sayarpour, M. and Yousef, A.A. 2007. Optimization of Oil Production Based on a Capacitance Model of Production and Injection Rates. Paper SPE 107713, presented at the 2007 SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, TX, 1-3 April.

Moncur, C.E., Jakeman, S.V.J., Berendschot, L. et al. 2008. Extensions to and Roll out of Data Driven Production Surveillance and Optimization. Paper presented at the Intelligent Energy Conference and Exhibition, Amsterdam, The Netherlands. Society of Petroleum Engineers 112037-MS.

Nobakht, M. and Mattar, L. 2009. Diagnostics of Data Quality for Analysis of Production Data. Paper presented at the Canadian International Petroleum Conference, Calgary, Alberta. Petroleum Society of Canada 2009-137.

Rubin, E., Rosenegger, L., and Mesmari, N. 1997. Approach to Handle Lost or Questionable Production Data. Paper presented at the Annual Technical Meeting, Calgary, Alberta. Petroleum Society of Canada 97-43.

Sayarpour, M., Zuluaga, E., Kabir, C.S. and Lake, L.W. 2007. The Use of Capacitance-Resistive Models for Rapid Estimation of Waterflood Performance and Optimization. Paper SPE 110081 presented at the SPE Annual Technical Conference and Exhibition, Anaheim, California, 11-14 November.

Sayarpour, M. 2008. Development and Application of Capacitance-Resistive Models in Water/CO2 Floods, Ph.D. Dissertation. The University of Texas at Austin, Austin, Texas.

Shipley, D., Weltevrede, B., Doniger, A. et al. 2008. Production Data Standards: The Prodml Business Case and Evolution. Paper presented at the Intelligent Energy Conference and Exhibition, Amsterdam, The Netherlands. Society of Petroleum Engineers 112259-MS.

Weber, D.B., Hales, H.B., and Baxter, L.L. 2004. A New Method of Formulating Finite Difference Equations – Some Reservoir Simulation Examples. Paper 2004-170 presented at the Canadian International Petroleum Conference, Calgary, Alberta, Canada, 8-10 June.

Weber, D.B., Edgar, T.F., Lake, L.W., Lasdon, L.S., Kawas, S., and Sayarpour, M. 2009. Improvements in Capacitance-Resistive Modeling and Optimization of Large Scale Reservoirs. Paper SPE 121299 presented at the SPE Western Regional Meeting, San Jose, California, 24-26 March.

Weber, D.B. 2009. The Use of Capacitance-Resistance Models to Optimize Injection Allocation and Well Location in Water Floods, Ph.D. Dissertation. University of Texas at Austin, Austin, Texas.

Wising, U., Vrielynck, B., Kalitventzeff, P.-B. et al. 2009. Improving Operations through Increased Accuracy of Production Data. Paper presented at the Offshore Europe, Aberdeen, UK. Society of Petroleum Engineers 124766-MS.

Yousef, A.A., Gentil, P.H., Jensen, J.L. and Lake, L.W. 2005. A Capacitance Model to Infer Interwell Connectivity from Production and Injection Rate Fluctuations. Paper SPE 95322 presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 9-12 October.

Yousef, A. A. 2005. Investigating Statistical Techniques to Infer Interwell Connectivity from Production and Injection Rate Fluctuations. Ph.D. Dissertation, The University of Texas at Austin, Austin, Texas.

Yousef, A.A., Gentil, P.H., Jensen, J.L. and Lake, L.W. 2006. A Capacitance Model to Infer Interwell Connectivity from Production and Injection Rate Fluctuations. SPEREE, 9 (5): 630-646.

Yousef, A.A., Jensen, J.L. and Lake, L.W. 2006. Analysis and Interpretation of Interwell Connectivity From Production and Injection Rate Fluctuations Using a Capacitance Model. Paper SPE 99998 presented at the SPE/DOE Symposium on Improved Oil Recovery, Tulsa, Oklahoma, 22-26 April.

Yousef, A.A., Jensen, J.L. and Lake, L.W. 2009. Integrated Interpretation of Interwell Connectivity Using Injection and Production Fluctuations. Mathematical Geosciences, 41 (1): 81-10